Computational investigation of intracellular signalling cascades: from G-protein-coupled receptors to adenylyl cyclase

THÈSE N^{O} 7563 (2017)

PRÉSENTÉE LE 22 FÉVRIER 2017
À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE CHIMIE ET BIOCHIMIE COMPUTATIONNELLES
PROGRAMME DOCTORAL EN CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Siri Camee VAN KEULEN

acceptée sur proposition du jury:

Prof. S. Gerber, présidente du jury
Prof. U. Röthlisberger, directrice de thèse
Dr S. Vanni, rapporteur
Prof. P. Carloni, rapporteur
Prof. M. Dal Peraro, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2017

Figure 0.1: Untitled. Siri Camee (2015).

Acknowledgements

First and foremost I want to thank Ursula Röthlisberger for being my supervisor and for supporting me during my doctoral studies to become a publishing scientist. Meetings with her provided me new insights in projects as well as a better understanding of how science works and how to progress in challenging studies. As a person, I highly respect the way she handles life, how she helps people in the process and how she shares knowledge with fellow scientists. Of course I cannot forget Ruby, Quanta and Topas who always provided lots of love and cuddles.

I would also like to thank Elisa Liberatore, whom I was very lucky to have shared an office with for the passed five years. Throughout those years we have gone through many stages in our lives. However, I have always felt that we were able to support one another in our scientific and personal endeavours. Of course I also have to thank her for the great job she did by translating the English abstract of this thesis into Italian.

One of my collaborators during the doctoral studies was Lucie Delemotte whom I would like to thank for inspiring me with her strength, enthusiasm and determination in science and in life.

I also want to thank Martin Peter Bircher and Thibaud von Erlach for all the nice board game evenings and tea breaks, which gave me the opportunity to laugh a lot, eat lots of sweets and make nice memories.

Negar Ashari Astani was one of my fellow PhD colleagues that just started before me in the LCBC group. She immediately made me feel like I was part of the group and we shared many fun moments inside and outside the lab.

In the PhD I had the pleasure to work together with many scientists. I would like to thank Alicia Solano, Daniele Narzi, Murat Kiliç, Pietro Vidossich, Vincenzo Carnevale, Eleonora Gianti, Przemyslaw Nogly and Jörg Standfuss for their collaboration on several projects, which I have experienced as very stimulating. Pablo Campomanes Ramos helped me during the starting period of my doctoral studies, which I very much appreciated. I also want to thank him for always being available for scientific discussions even after he left the lab. I would like to thank Manuel Dömer for helping out whenever questions surfaced about force matching.

During my time in the LCBC group I also enjoyed working with Karin Pasche, who was always there to help and with whom it was a blast to organise christmas dinners for the group.

Of course I would also like to thank Ariadni Boziki, Giulia Palermo, Marta Andreia Da Silva

Perez, Pablo Lopez Tarifa, Polydefkis Diamantis, Basile Curchod, Bruno Horta, Elizabeth Brunk, Lorenzo Casalino, Felipe Franco de Carvalho, Tom Penfold, Swarnendu Bhattacharyya, Esra Bozkurt, Nicholas John Browning and all the other members of the LCBC laboratory for the nice time in the lab.

Part of my doctoral studies was embedded in the Human Brain Project (HBP). I would like to thank all the members of the HBP subgroup, Jeanette Hällgren Kotaleski, Paolo Carloni, Rebecca Wade, Richard Lavery, Neil Bruce, Juliette Martin, Elisa Frezza and Nicoletta Ceres, for their collaboration and stimulating scientific discussions.

I would also like to thank Sandrine for being the jury president during the private defense and the committee members of the jury, Paolo Carloni, Matteo dal Peraro and Stefano Vanni for their interesting questions and scientific discussion.

Before joining the LCBC group I performed my master project at the University of Amsterdam in the group of Peter Bolhuis, which motivated me to pursue a PhD. I would like to thank Peter Bolhuis, Bernd Ensing, Jocelyne Vreede, Evert Jan Meijer and Marieke Schor for their support and for introducing me to computational biochemistry.

Besides many colleagues who supported me during my doctoral studies, I have had a lot of support from my family as well. Although I know it was very hard for my parents that I did not live close by, they were always there for me to talk to, to comfort, to care and to love. I want to truly thank my parents, Ine and Sybrandt, for supporting me and for stimulating me to perform the best I can. Mom, dad, I love you dearly and I am proud of you being my parents. I also want to thank my dear sister, Siepke, who was always there for me to share the good times and the bad times. I love you sis! Her long stays in Switzerland made me feel like she brought a bit of home from Amsterdam to Ecublens.

I also want to thank Kasper's parents and sister, Ina, Tjalling and Ariëlle, for being very welcoming and for helping wherever they can. During the passed five years of my doctoral studies I lived together with my sweetheart Kasper. I want to thank him for his full support, his love, his care and for always being there whenever necessary, which helped me to put as much time as needed into my studies and to have fun in the process.

Abstract

In order to adjust to changes in the environment, cells can communicate via signalling cascades. A common pathway for intercellular communication is the G-protein-coupled-receptor (GPCR) signal transduction cascade which is triggered by a (chemical or physical) extracellular signal, such as light or small molecules. This external interaction induces conformational changes that can lead either to GPCR activation or deactivation. In the active form, GPCRs can induce the activation of trimeric G proteins, which results in the dissociation of the Gα from the $\beta\gamma$ subunits. Subsequently, active Gα can interact with other cytosolic proteins, such as adenylyl cyclase (AC), which converts adenosine triphosphate to cyclic adenosine monophosphate (cAMP) upon stimulation. AC activation can be induced via stimulatory Gα subunit (Gα_s) interactions, while inhibition is related to inhibitory Gα (Gα_i) association.

X-ray crystallography is crucial for the computational study of the GPCR signal transduction pathway. However, sometimes important molecules or moieties can become lost or cannot be incorporated during crystallisation procedures, which can have a significant effect on a protein's conformation and function. In this thesis, three steps in the GPCR signal transduction cascade have been studied, rhodopsin activation (i), the conformation of active Gα_i (ii) and the interaction between Gα_i and AC (iii). In all three cases some groups or molecules that are absent in the X-ray structures play a vital role in the function of the proteins.

In the first project, rhodopsin's early intermediates after photon exposure are investigated together with a deprotonation reaction that takes place in the later stages of the activation pathway. The classical molecular dynamics (MD) and mixed quantum mechanics/molecular mechanics MD results suggest that photon exposure induces active site rearrangements which increase the stability of rhodopsin's deprotonated Schiff base state. A bridging water molecule in the active site that is absent in the X-ray structure appears to play an important role during the deprotonation mechanism, which demonstrates the dependence of the protein's function on its environment.

The second study investigates the conformation of the active soluble form of Gα_i, which has not yet been fully resolved via experimental studies due to crystallisation difficulties of the lipidated N-terminus. To investigate the effect of lipidation via classical MD simulations, a model of soluble lipid-bound Gα_i was constructed. The simulations show that Gα_i can form a hydrophobic pocket for the lipid on the protein surface. Other regions of the protein, important for protein-protein interaction, are adjusted as well. Hence, the post-translational

modification appears to have a significant impact on Gα_i's active conformation and function, which was not captured via the unlipidated X-ray structures.
The last project is devoted to a complex of active soluble Gα_i and AC (Gα_i:AC). While the interaction site for Gα_s on AC is known, the interaction site for Gα_i has not yet been identified. Classical MD simulations were performed on a docked Gα_i:AC complex to understand the inhibition mechanisms as well as the differentiation between AC stimulation/inhibition. The results show that Gα_i binding not only leads to AC inhibition via impeding ATP binding but also by preventing re-activation via Gα_s through closure of the subunit's interaction site on AC.

Keywords: GPCR signal transduction pathway, G-protein-coupled recepter, rhodopsin, G protein, complex formation, cis-trans isomerisation, adenylyl cyclase

Riassunto

Per adattarsi ai cambiamenti dell'ambiente, le cellule possono comunicare attraverso segnali a cascata. Un meccanismo tipico di comunicazione è la cascata di trasduzione del segnale dei recettori accoppiati a proteine G (GPCR), innescata da un segnale extracellulare chimico (piccole molecole) o fisico (fotoni). Lo stimolo esterno induce cambiamenti conformazionali che possono portare all'attivazione o alla disattivazione dei GPCR. GPCR attivi possono indurre l'attivazione delle proteine G trimeriche, che porta alla dissociazione delle sub-unità Gα da quelle $\beta\gamma$. In seguito, le Gα attive possono interagire con altre proteine citosoliche, come la adelinilico ciclasi (AC), che converte l'adenosina trifosfato in adenosina monofosfato ciclica (cAMP), dopo la stimolazione. L'attivazione delle AC può essere indotta da interazioni con Gα stimolatrici (Gα_s); l'inibizione è legata all'associazione di Gα inibitorie (Gα_i).

La cristallografia a raggi X è fondamentale nello studio dei meccanismi di trasduzione. Tuttavia, alcune molecole o parti di esse, possono non venire incorporate durante la cristallizzazione, con un impatto anche notevole sulla conformazione o sulla funzione della proteina. In questa tesi sono stati studiati tre aspetti della cascata di trasduzione del segnale dei GPCR: l'attivazione della rodopsina (i), la conformazione della Gα_i attiva (ii) e l'interazione tra Gα_i e AC (iii). In tutti i casi, alcuni gruppi funzionali o molecole non presenti nelle strutture cristallografiche sono cruciali per la funzione delle proteine.

Nel primo progetto sono stati studiati sia gli intermedi iniziali della rodopsina, seguenti l'esposizione a fotoni, che la reazione di deprotonazione che avviene in una fase posteriore. I risultati di simulazioni di dinamica molecolare (MD) classica e mista quantomeccanica/meccanica molecolare (QM/MM) suggeriscono che la foto-eccitazione induce una riorganizzazione del sito attivo che aumenta la stabilità dello stato deprotonato della base di Schiff della rodopsina. Una molecola d'acqua "pont" nello stato attivo, assente nelle strutture da raggi X, sembra essere importante nel meccanismo di deprotonazione: ciò dimostra la dipendenza della funzione della proteina dall'ambiente ad essa circostante.

Il secondo studio riguarda la conformazione della forma attiva solubile della Gα_i, non ancora completamente determinata sperimentalmente per via di difficoltà di cristallizzazione del dominio N-terminale lipidato. Per indagare l'effetto della lipidazione è stato costruito un modello di Gα_i solubile legata a un lipide. Simulazioni MD classiche mostrano che la Gα_i può formare una tasca idrofobica per il lipide sulla superficie della proteina. Anche altre

importanti regioni della proteina risultano modificate: la modificazione post-traduzionale sembra avere un ruolo significativo sulla conformazione e sulla funzione della Gα_i, ruolo non colto attraverso strutture non-lipidate.
L'ultimo progetto è dedicato a un complesso dell'unità attiva solubile Gα_i e dell'AC, (Gα_i:AC). Il sito di interazione della Gα_i sulla AC non è stato ancora identificato. Simulazioni MD classiche di un complesso accoppiato (Gα_i:AC), effettuate per capire i meccanismi di inibizione e le differenziazioni tra stimolazione e inibizione dell'AC, mostrano che il legame con la Gα_i porta all'inibizione dell'AC non solo impedendo il legame ATP ma anche prevenendo la riattivazione via Gα_s, mediante chiusura del sito di interazione della sub-unità sulla AC.

Parole chiave: meccanismo di trasduzione del segnale nelle GPCR, recettori accoppiati a proteine G, adelinilico ciclasi, formazione di complessi, isomerizzazione cis-trans.

Contents

List of Figures

List of Tables

1 Introduction

In everyday life, medication is used as a remedy for a vast amount of diseases. Drugs can help to relieve symptoms such as pain but are also used in the fight against Parkinson's and heart disease, for example. Close to 40% of drugs on the market target the G-protein-coupled-receptor (GPCR) signal transduction pathway as this pathway is widely used for intercellular signalling within the human body. Most GPCR based signalling pathways are triggered by ligands that are too large to directly travel across the membrane and interact therefore with membrane-embedded GPCRs to transfer the signal to the intracellular region. Besides drugs, examples of natural ligands are hormones, neurotransmitters, small proteins but also chemical or physical exogenous signals such as caffeine or photons. Ligands can function as agonists, antagonists or inverse agonists, depending on their specific interaction with a GPCR. Once the signal has arrived in the cytosol via a GPCR, activated by an agonist, several proteins have to be employed in order to achieve the final biological result, which can range from cell growth to apoptosis. In order to get a better understanding of the way an extracellular signal impacts the conformation of a GPCR as well as the transmission of the extracellular signal into the cytosol, three steps in the GPCR signal transduction pathway have been investigated in this thesis, namely rhodopsin activation, the conformation of active inhibitory G proteins and their interaction with the enzyme adenylyl cyclase. The proteins that play an important role in these steps (rhodopsin, G proteins and adenylyl cyclase) are described in more detail in Section 1.1. The last sections of the introduction contain the problem statement and the outline of this thesis.

1.1 Context and Motivation

1.1.1 Rhodopsin, a Prototypical G-Protein-Coupled Receptor

The G-protein-coupled receptor family includes around 800 GPCRs [Lagerström and Schiöth, 2008], which are present in, for instance, the pulmonary system, the central nervous system, the immune system, the eye and muscles [Borea et al., 2015]. GPCRs have a large impact on the function of the human body because they are located in the membrane of cells which enables them to relay communication between the intracellular and extracellular side of a cell. Active GPCRs are able to stimulate or inhibit several proteins, which include protein kinase A (PKA) and C (PKC) as well as RhoA (Fig. 1.1a). The activation of these proteins can have a range of effects in the cell. PKA, for example, when activated in fat cells, can stimulate the breakdown of fat [Patrick, 2013].

GPCRs are divided into six classes, from class A to class F, that share similarities in sequence and function. The largest class, class A, includes around 700 GPCRs that are composed of seven trans-membrane domains, three intracellular loops, three extracellular loops, a C-terminus and an N-terminal domain (Fig. 1.1b). One of class A's GPCRs rhodopsin is located in the outer-segment of rod cells in the eye (Fig. 1.2b). Besides the structural elements described in figure 1.1, rhodopsin also includes an 11-*cis*-retinal moiety that is covalently linked to the

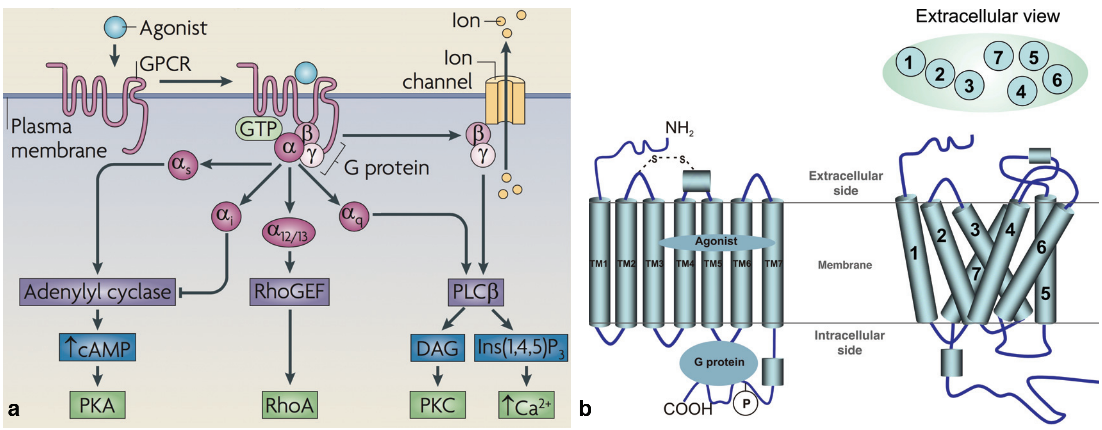


Figure 1.1: Description of the G-protein-coupled-receptor (GPCR) signal transduction cascade and a class A GPCR template structure. (a) Activation of intracellular pathways via different G-protein families, induced via GPCR stimulation, leads to the activation or inhibition of adenylyl cyclase, phospholipase Cβ or RhoGEF, which ultimately result in the activation or inhibition of protein kinase A, RhoA, protein kinase C, or a concentration increase in Ca^{2+}. Taken from Ritter and Hall [2009]. (b) Template class A GPCR structure showing the G-protein binding site and the general ligand binding site together with the trans-membrane domains and the N-terminal and C-terminal tail. Taken from Foreman et al. [2010].

receptor in the active site (Fig. 1.2a, c) [Smith, 2010]. Before rhodopsin photoactivation, retinal acts as an inverse agonist in the active site, however, after light exposure, the chromophore is converted to the GPCR's agonist [Struts et al., 2011]. This change from inverse agonist to agonist is induced by the ultrafast photoisomerisation of an 11-*cis* to an all-*trans* configuration of the chromophore (Fig. 1.2a) which triggers a series of events that start near the retinal moiety in the active site and ultimately spreads over the entire protein (Fig. 1.2d).

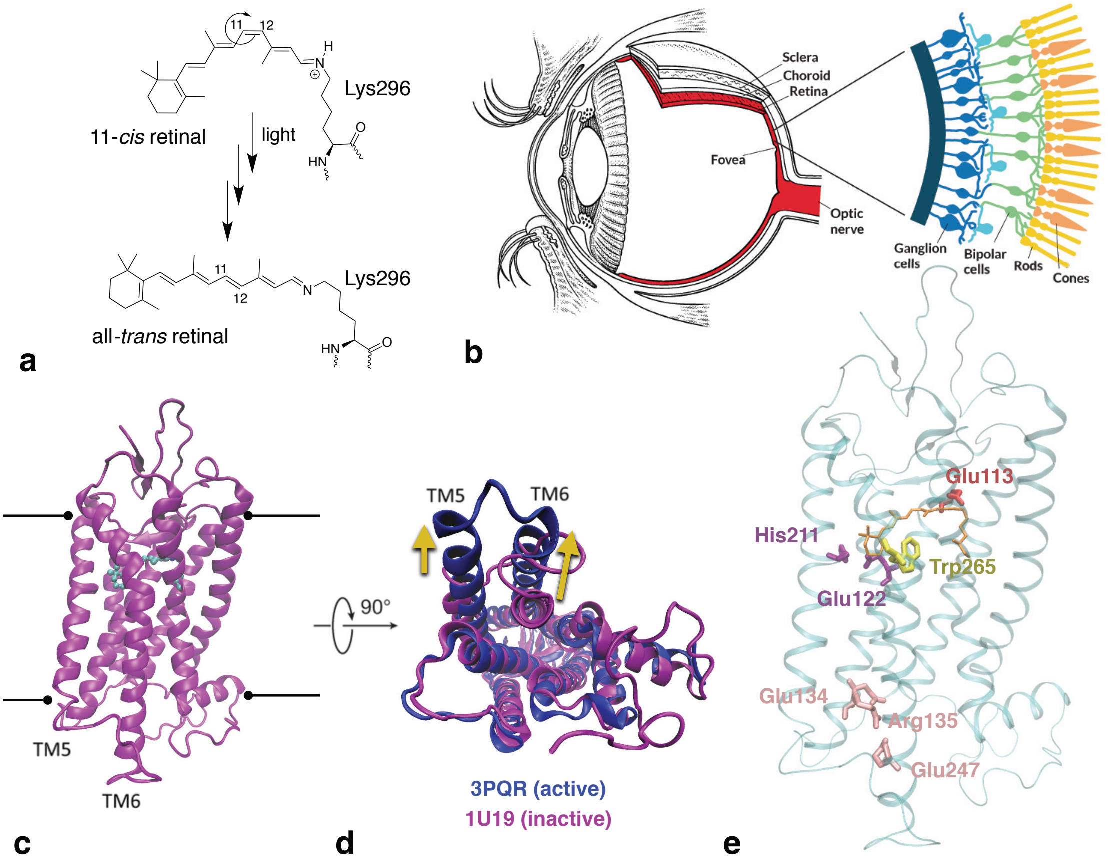


Figure 1.2: Rhodopsin's activation mechanism. (a) 11-*cis* to all-*trans* photoisomerisation and later deprotonation of covalently bound retinal in rhodopsin due to exposure to light. (b) Schematic representation of the human eye and the composition of the retina, showing the rods, in which rhodopsin is located, and the cone cells. Taken from Saey [2015]. (c) X-ray structure of inactive rhodopsin (PDB code 1U19). The black lines indicate the location of the cell membrane. (d) Alignment of the active and inactive conformation of rhodopsin, viewed from the intracellular side. Corresponding PDB codes are listed in the same colour as the colour of the protein structure. The main structural changes due to activation are indicated by yellow arrows. (e) X-ray structure of inactive rhodopsin (PDB code 1U19) in which specific regions that are especially affected by retinal's *cis*-*trans* isomerisation are highlighted. The retinal moiety is shown in orange.

After the chromophore's change in configuration, also its orientation in the active site is gradually altered. The subsequent relaxation induces deprotonation of the protonated Schiff base (PSB) of Lys296, covalently linked to the retinal group, and protonation of the counter ion Glu113 (Fig. 1.2a, e). This deprotonation mechanism is highly influenced by rhodopsin's environment since, for example, deprotonation takes place earlier in the activation pathway at room temperature than when rhodopsin is exposed to photons at low temperatures [Sandberg et al., 2014; Thorgeirsson et al., 1993]. The pathway leading to the active conformation of rhodopsin also includes more intermediate states at room temperature than when low temperatures are applied to the system, as, for instance, during X-ray crystallisation. Besides PSB deprotonation, a change in the position of Trp265 takes place in the photoisomerised all-*trans* retinal configuration as well as a rearrangement of the hydrogen-bond network around Glu122 and His211 (Fig. 1.2e). Further away from the active site in the intracellular region, the isomerisation ultimately affects the "closed" interaction site for G proteins that includes the conserved residues Glu134, Arg135 and Glu247, called the ionic lock (Fig. 1.2e). As part of the activation process, the salt bridges Glu134-Arg135 and Arg135-Glu247 are broken and Glu134 is protonated, leading to an "open" conformation of the intracellular region, the signalling state meta II (Fig. 1.2c, d, e). Rhodopsin's meta II state is able to interact with G proteins in the cytosol that will augment the signal as one active rhodopsin can stimulate multiple G proteins. Ultimately, rhodopsin induces a signal from the retina to the brain which results into eye sight for dim light.

1.1.2 Down-Stream Effect: G Proteins and Adenylyl Cyclase

One of the main interaction partners of active GPCRs on the cell's intracellular side are G proteins (Fig. 1.1a and Fig. 1.3a). G proteins consist of three subunits: α, β and γ. While the GPCR family contains around 800 different GPCRs, the human body includes only a relatively small variety of G-protein subunits: 21 α, 6 β and 12 γ subunits [Oldham and Hamm, 2008]. When a GPCR is activated via the interaction with an agonist, G proteins can start interacting with the GPCR's intracellular region via the C-terminus of the Gα subunit (Fig. 1.3a). Due to the interaction with the active GPCR, the G protein is able to substitute guanosine diphosphate (GDP) for guanosine-5'-triphosphate (GTP), located at the interface between Gα and the β subunit. After the GDP/GTP exchange, Gα dissociates from the $\beta\gamma$ subunits and can transmit the extracellular signal by interacting with other proteins in the cytosol, such as adenylyl cyclase (AC) or phospholipase Cβ (Fig. 1.1a). An active Gα protein can stimulate or inhibit enzymes in the cell, depending on which subfamily the subunit belongs to: Gα_s, Gα_i, Gα_q or Gα_{12}.

While the Gα_q and Gα_{12} subfamilies activate RhoA or PKC, Gα_s and Gα_i both affect the same enzyme, AC, in an inverse fashion. Ten isoforms of AC are known of which nine are membrane-bound (AC1-9) and one is soluble (sAC). Gα_s can stimulate all membrane-bound isoforms of AC, leading to the conversion of adenosine triphosphate (ATP) to 3',5'-cyclic adenosine monophosphate (cAMP) and pyrophosphate (Fig. 1.3b), while Gα_i can only inhibit isoforms AC1, AC5 and AC6 [Sadana and Dessauer, 2008; Sunahara et al., 1996]. Another striking difference between the two Gα subunits is the lipidation of the N-terminus. Both Gα subunits undergo palmitoylation, the addition of a saturated 16-carbon fatty acid to the Gα subunit

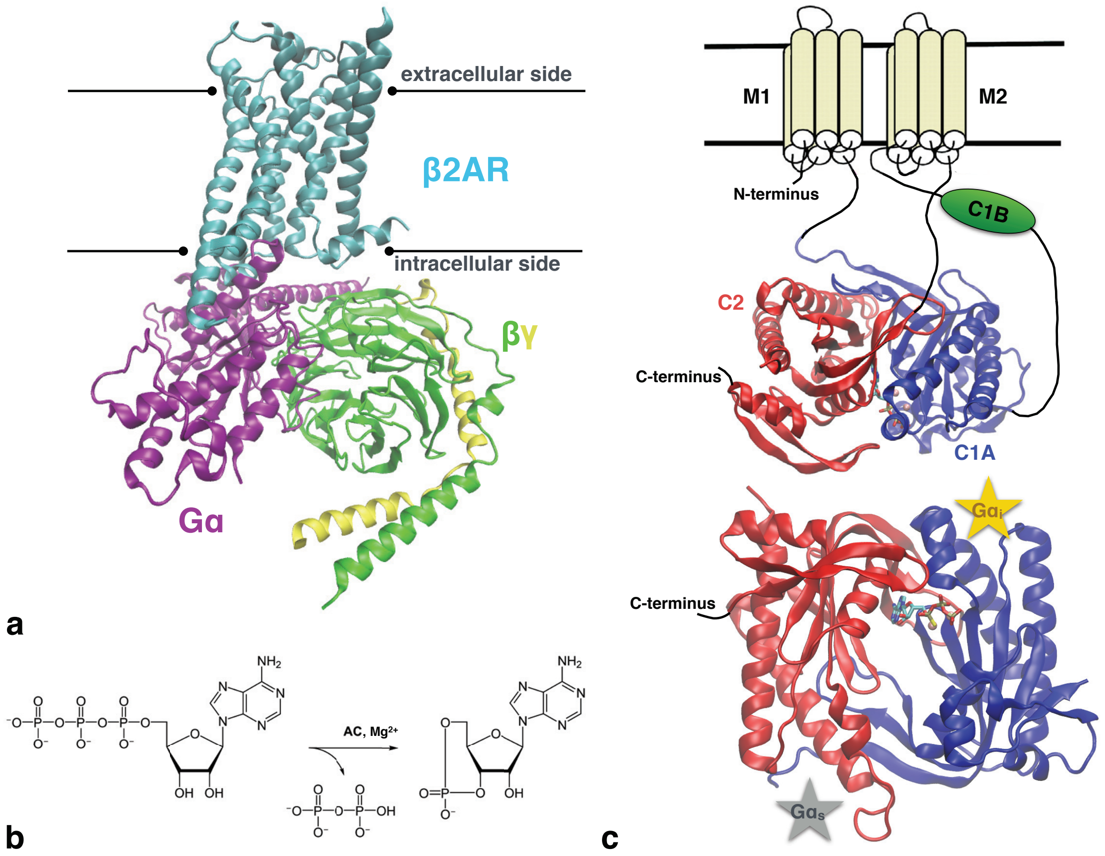


Figure 1.3: Proteins on a cell's intracellular side that can become activated or inhibited in response to an extracellular signal, transmitted via GPCRs. (a) X-ray structure of $\beta2$ adrenergic receptor:G$_s$-protein complex (PDB code 3SN6) [Rasmussen et al., 2011]. The black lines indicate the location of the cell membrane. (b) Reaction mechanism that takes place in AC when the enzyme is stimulated via active Gα_s. In this reaction, adenosine triphosphate (ATP) is converted to 3',5'-cyclic adenosine monophosphate (cAMP) and pyrophosphate. Taken from Wikipedia. (c) Schematic representation of the complete membrane-bound AC5 protein in the top panel, showing the unresolved protein regions, the trans-membrane domains and the CB1 domain via cartoon representation. PDB code 1CJK was used to describe the catalytic domains, C1A and C2. In the bottom panel the catalytic domains, C1A and C2, and their pseudo-symmetrical interaction sites for Gα_s and Gα_i together with the location of an ATP analog in the active site are depicted. The images were inspired by Hurley [1999].

via esterification of a cysteine residue by a palmitate molecule [Cabrera-Vera et al., 2003], which is a reversible post-translational modification. However, only Gα_i is also subjected to an irreversible addition of a saturated 14-carbon fatty acid, a myristoyl moiety, to Gly2 via an amide bond. Gly2 becomes the first residue of the Gα_i sequence, after Met1 is removed by methionine amino-peptidase [Cabrera-Vera et al., 2003]. The role of lipidation has not been fully understood yet but it is known that palmitoylation is required to stably anchor Gα subunits to the membrane [Degtyarev et al., 1994]. Solely myristoylated Gα_i subunits shift from being localised at the membrane to being both present at the membrane as well as in the cytosol [Degtyarev et al., 1994]. In the case of Gα_i, myristoylation impacts not only the subunit's location in the cell but also its active conformation and its inhibitory function. Non-myristoylated Gα_i is unable to inhibit AC as the Gα subunit cannot form a complex with the enzyme anymore [Preininger et al., 2003; Taussig et al., 1993].

Gα_s's active conformation and its stimulatory interaction with AC are well known as several X-ray structures exist of active Gα_s structures as well as of a Gα_s subunit interacting with AC's catalytic C2 domain (Fig. 1.3c) [Sunahara et al., 1997a,b]. The Gα_s X-ray structures co-crystallised with AC do not contain the trans-membrane domains and the C1B domain but this pruned version of AC is still able to convert ATP when the enzyme is in its stimulated form [Dessauer et al., 1998]. In contrast to the large number of AC:Gα_s complexes and active Gα_s structures, no Gα_i:AC complex or myristoylated forms of active Gα_i are present in the protein data bank (PDB). Although non-myristoylated active Gα_i X-ray structures are available, this form of Gα_i cannot inhibit AC activation and is therefore not an appropriate candidate for comparison with active Gα_s structures [Preininger et al., 2003]. The only currently available experimental study that has investigated myristoylated Gα_i's binding site on AC is a mutagenesis study by Dessauer et al. [1998]. This study indicates that Gα_i binds to the pseudo-symmetrical binding site on the C1A domain (Fig. 1.3c), which will be called C1 domain throughout this thesis. Hence, structural understanding of how Gα_s and Gα_i bias AC into an active or inactive conformation is still unclear.

1.2 Problem: Activation and Regulation

In cell communication, activation as well as its regulation are important. The signal transmission of active GPCRs to the cytosol can either lead to the stimulation or the inhibition of a pathway. The GPCR rhodopsin is able to stimulate the brain via photons, resulting in vision when only dim light is available. After photon exposure, the transfer of retinal's configurational change to the protein environment is generally well understood. However, one of the questions that remain is: What enables or triggers proton transfer and why is it temperature dependent?

When moving further downstream in the GPCR signal transduction cascade, AC regulation can take place in the cytosol as GPCRs can activate stimulatory and inhibitory G proteins. The conformation of Gα_s subunits in their active state and their interaction site on AC are known from X-ray structures. However, the active conformation of myristoylated Gα_i is still under debate as the available X-ray structures of the Gα_i subunit do not incorporate N-terminal myristoylation. The absence of the myristoyl moiety results in a significant change in Gα_i structure as well as an incapability to inhibit AC since no Gα_i:AC complex can be formed. Because no X-ray structures of the Gα_i:AC complex have been resolved, the interaction between myristoylated Gα_i and AC is not well understood. Although mutagenesis studies are available, the results are inconclusive about Gα_i's specific interaction site. Hence, due to the uncertainty of the effect of myristoylation and the absence of X-ray crystallographic data, Gα_i's interaction with AC and its mechanism of inhibition as well as the difference in action of Gα_s versus Gα_i in spite of their structural similarity have not yet been elucidated.

The study of the inhibitory and stimulatory mechanism that Gα subunits exert on AC was embedded in the Human Brain project (HBP). In the subgroup within HBP, European laboratories, specialised in systems biology, Brownian dynamics, classical MD and QM/MM MD simulations, collaborated to obtain a better understanding of AC regulation via G proteins. The study mainly focussed on the stability of possible complexes AC can form with Gα_s and Gα_i as well as each Gα's effect on AC's function.

1.3 Organisation of this Thesis

After a short summary of the underlying theory of the computational methods used in this thesis in Chapter 2, Chapters 3-5 are presenting the obtained results for the three projects. The first step of the GPCR signal transduction cascade described in Chapter 3, is the activation of rhodopsin. This study focusses on the early intermediates of the protein after photon exposure and the PSB deprotonation reaction that takes place in the active site. In this project classical MD and QM/MM MD simulations have been applied.

The next step that follows in the signal transduction pathway is the association of a G protein to the intracellular region of a GPCR, which leads to the activation of a Gα subunit. This active conformation of Gα is covered in Chapter 4 in which the active myristoylated Gα_i configuration is studied, located in the cytosol. This investigation was conducted via classical MD of two different configurations of the Gα_i subunit, with and without the myristoylated N-terminus.

In Chapter 5, the interaction between the newly determined Gα_i conformation and AC is explored by comparing classical MD simulations of AC interacting with Gα_i as well as an AC protein by itself.

The sixth and final chapter summarises the findings of Chapters 3-5 and discusses future perspectives of the performed exploratory simulations.

2 Theory

In this thesis a variety of computational approaches has been used to investigate the G-protein-coupled receptor (GPCR) rhodopsin, the inhibitory G protein (G_i) and the enzyme adenylyl cyclase (AC). This chapter gives a short summary of the theoretical foundations of the different methods.

The majority of the following sections is based on the textbooks by Atkins, Leach and Frenkel [Frenkel and Smit, 2002; Atkins and Friedman, 2004a; Leach, 2001]. In this chapter the molecular dynamics (MD) method (Section 2.1) is described in the form of *ab initio* MD (QM MD) (Section 2.2) and force-field based MD (MM MD) (Section 2.3). The combination of QM and MM MD is discussed in Section 2.4. The last section of Chapter 2, Section 2.5, covers the force matching method, which is used for the automatised parameterisation of MM force fields from QM MD data.

2.1 Molecular Dynamics

MD is a widely used computational technique to monitor the classical movement of particles [Frenkel and Smit, 2002]. To describe the time evolution, Newton's equations of motion,

$$\mathbf{f}_i = m_i \mathbf{a}_i,$$

$$\mathbf{f}_i = -\frac{dV}{d\mathbf{r}_i}, \qquad \mathbf{a}_i = \frac{d\mathbf{v}_i}{dt}, \qquad \mathbf{v}_i = \frac{d\mathbf{r}_i}{dt},$$

have to be integrated in time. In these equations $\mathbf{f}$ is the force, $\mathbf{a}$ is the acceleration, $\mathbf{v}$ is the velocity, V is the total potential energy, $\mathbf{r}$ is the position of the particle and i is the index for particle i. Integration in time is usually achieved by using numerical integration in discrete time steps Δt. During an MD simulation the positions and velocities are calculated for each time step, Δt, – a very short period of time for which it is assumed that the forces on the particles are constant– generating a time series $t_0 + \Delta t$, $t_0 + 2\Delta t$,...etc. If the force is not constant within one time step, the calculation of the equations of motion becomes erroneous. Hence, it is of great importance to specify a sufficiently small time step that generates a minimum of force divergence during each time step and is still efficient. Two widely used methods for calculating the force on particles and their velocities in MD simulations are the Velocity Verlet and the Leapfrog algorithm [Frenkel and Smit, 2002]. These methods have both been applied during this thesis since the QM MD as implemented in the CPMD code uses the Velocity Verlet algorithm and the MM MD as implemented in the GROMACS package employs the Leapfrog technique [Hess et al., 2008; CPMD].

The initial positions of a system are setup before the start of a simulation using a structure from the protein data bank, for example, and the initial velocities have to be generated e.g. from a Maxwell-Boltzmann distribution at a given temperature.

Newton's equations of motion can also be derived from the Euler-Lagrange formula,

$$\frac{d}{dt}\frac{\delta L}{\delta \dot{\mathbf{r}}_i} - \frac{\delta L}{\delta \mathbf{r}_i} = 0, \tag{2.1}$$

with the classical Lagrangian,

$$L(\mathbf{r}^N, \dot{\mathbf{r}}^N) = \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\mathrm{m_i}}{2}\dot{\mathbf{r}}_{\mathrm{i}}^2 - \mathrm{V}(\mathbf{r}^{\mathrm{N}}), \tag{2.2}$$

with m_i being the mass of atom i, $\mathbf{r}^{\mathrm{N}}$ being the generalised coordinates and $\dot{\mathbf{r}}^{\mathrm{N}}$ the time derivatives of the coordinates. The Lagrangian, $L(\mathbf{r}^N, \dot{\mathbf{r}}^N)$, in equation 2.2, which is especially suitable for adding constraints to the dynamics of the system, is defined as the difference between the kinetic energy and the potential energy, $V(\mathbf{r}^{\mathrm{N}})$.

As previously discussed, the force on the particles is essential to calculate the system's motion. However, the total potential, V (Eq. 2.1), that is required to obtain the forces for each time step cannot be determined in an exact way for many-body quantum systems of more than three particles. Consequently, an approximate approach is required to obtain the forces on the particles. There are different methods to obtain an approximate potential and the choice for a specific approach depends on the size of the system and the events under study. If bond-breaking events occur, a quantum mechanical method has to be applied in order to accurately calculate the particle forces in the system (see Section 2.2). Yet, at first-principles level this is only feasible to a maximum of several hundreds to thousands of atoms and is in any case computationally demanding. Molecular mechanics or classical MD (see Section 2.3), based on parameterised force fields with limited transferability, is computationally more feasible for large systems of several hundred thousands of atoms.

2.2 Ab Initio Molecular Dynamics

To study a reaction mechanism, a quantum mechanical MD method has to be applied to explicitly describe the changes in the electron density. The time evolution of a quantum system is given by the time-dependent Schrödinger equation,

$$\hat{H}(\mathbf{r}^N, \mathbf{R}^M, t)\Phi(\mathbf{r}^N, \mathbf{R}^M, t) = i\hbar\frac{\delta}{\delta t}\Phi(\mathbf{r}^N, \mathbf{R}^M, t), \tag{2.3}$$

where $\hat{H}$ is the Hamiltonian, Φ is the wavefunction for the electrons and the nuclei, $\mathbf{r}^{\mathrm{N}}$ are the set of spatial coordinates of N electrons, $\mathbf{R}^{\mathrm{M}}$ are the set of spatial coordinates of all M nuclear coordinates, t is time and $\hbar$ is the reduced Planck constant. In this representation the electrons and nuclei are taken simultaneously into account, which results in complex calculations that are not computationally feasible for the simulation of larger systems.

One key simplification to solve this problem is the Born-Oppenheimer (BO) approximation [Atkins and Friedman, 2004a]. This approximation relies on the difference in masses between the nuclei and electrons which results in the fact that the motion of electrons is significantly faster than that of the nuclei [Atkins and Friedman, 2004a]. Because of this difference in characteristic time scales, the Hamiltonian $\hat{H}$ can be separated for nuclei and electrons and the electronic Schrödinger equation can be solved in the presence of fixed nuclear positions. Consequently, a second assumption of the BO approximation is that the nuclear dynamics follows a single potential energy surface. For regions, where the coupling between different electronic states becomes significant, the BO approximation breaks down and molecular dynamics approaches that include non-adiabatic effects [Tully, 1990; Thompson, 1998; Doltsinis and Marx, 2002] have to be applied. In the case of fully adiabatic dynamics, the potential energy surface can be determined from a solution of the time-independent non-relativist electronic Schrödinger equation:

$$\hat{H}_{ele}\Psi_{ele}(\mathbf{r}^N, \mathbf{R}_f^M) = E_{ele}\Psi_{ele}(\mathbf{r}^N, \mathbf{R}_f^M), \tag{2.4}$$

with $\hat{H}_{ele}$ being the electronic Hamiltonian, Ψ_{ele} the electronic wavefunction, E_{ele} the total electronic energy and $\mathbf{R}_f^M$ are the set of spatial coordinates of the fixed nuclei.

There are several first-principles techniques that are able to approximate E_{ele}, such as the Hartree-Fock method or Density Functional Theory (DFT). These techniques each provide their own ansatz for the electronic wavefunction, respectively the electronic density, and only need fundamental constants and atomic numbers to solve the time-independent electronic Schrödinger equation (Eq. 2.4) using the electronic Hamiltonian [Atkins and Friedman, 2004b] (in atomic units),

$$\begin{aligned}\hat{H}_{ele} &= \hat{T}_{ele} + \hat{V}_{nuc-ele} + \hat{V}_{ele-ele} \\ &= -\frac{1}{2}\sum_i \nabla_i^2 - \sum_{i,I} \frac{Z_I}{\mid \mathbf{r}_i^N - \mathbf{R}_I^M \mid} + \frac{1}{2}\sum_{i\neq j} \frac{1}{\mid \mathbf{r}_i^N - \mathbf{r}_j^N \mid}.\end{aligned} \tag{2.5}$$

2.2.1 Density Functional Theory

The electronic structure method used during this thesis work is Density Functional Theory (DFT). DFT is based on two theorems of which the first one states that the ground-state electron density determines a unique external potential (up to an additive constant) and therefore a unique H_{ele} and ψ. Consequently, all properties, including the total energy of the system, can be expressed as functionals of the electron density. The second theorem asserts that the minimisation of E_{ele} over all density distributions results in the ground-state energy and the ground-state density. Formal proofs of these statements were delivered by Hohenberg and Kohn in 1964 [Hohenberg and Kohn, 1964]. They determined an exact description of the ground-state electronic energy as a functional of ρ,

$$\begin{aligned}E_{ele} &= \langle \psi_0(\mathbf{R}_f) \mid \hat{H}_{ele} \mid \psi_0(\mathbf{R}_f)\rangle \\ &= E^{HK}[\rho_0] + \int d^3r\, V_{nuc-ele}(\mathbf{r})\rho_0(\mathbf{r}),\end{aligned} \tag{2.6}$$

in which $\mathbf{r}$ is a position vector in Euclidean space [Meijer, 2007]. $V_{nuc-ele}$ describes the Coulomb potential that arises from the nuclei and is usually called the external potential, V_{ext}, ρ_0 stands for the exact ground-state density and E^{HK} is a universal functional of the electron density. Although E^{HK} is exact and proven to exist, the specific form of the dependence on the electron density is not stated. In 1965, Kohn and Sham provided a reformation for this problem that has been used ever since [Kohn and Sham, 1965]. They defined $\rho(\mathbf{r})$ as a summation over a number of occupied single-electron orbitals, ψ_i, that describes a non-interacting single-particle system with the same density as the interacting system:

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} \mid \psi_i(\mathbf{r}) \mid^2. \tag{2.7}$$

Using the ψ_i orbitals (Eq. 2.7), denoted Kohn-Sham orbitals, the kinetic energy of a non-interacting system can easily be calculated, leading to single-particle equations to solve the Schrödinger equation, called the Kohn-Sham equations:

$$\epsilon_i\psi_i(\mathbf{r}) = H^{KS}\psi_i(\mathbf{r}) = [-\frac{1}{2}\nabla^2 + V_{eff}[\rho]]\psi_i(\mathbf{r}), \tag{2.8}$$

with the first term on the right defining the Kohn-Sham kinetic energy of the non-interacting single-particle system. $V_{eff}[\rho]$ includes the external potential $V_{ext}(\mathbf{r})$, the classical Coulomb potential and the exchange-correlation potential $V_{xc}[\rho]$,

$$V_{eff}[\rho] = V_{ext}(\mathbf{r}) + \int \mathrm{d}^3\mathrm{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \mathrm{V}_{\mathrm{xc}}[\rho]. \tag{2.9}$$

The definitions of $V_{ext}(\mathbf{r})$ and $V_{xc}[\rho]$ are:

$$V_{ext}(\mathbf{r}) = V_{nuc-ele}(\mathbf{r}) = -\sum_{I=1}^{M} \frac{Z_I}{|\mathbf{r}-\mathbf{R}_I|}, \tag{2.10}$$

$$V_{xc}[\rho] = \frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})}. \tag{2.11}$$

The exchange-correlation energy, E_{xc}, in equation 2.11 is a functional of the density as well and includes all non-classical electron-electron interactions, i.e. takes all many-body effects into account that are not incorporated into the other functionals [Atkins and Friedman, 2004b; Marx and Hutter, 2000]. However, the exchange-correlation energy is the only term that has not been expressed exactly so an approximation is necessary to determine the Kohn-Sham orbitals and thus the ground-state density. In most quantum mechanical calculations, the Kohn-Sham orbitals are expanded into an appropriate basis set:

$$\psi_i(\mathbf{r}) = \sum_k c_k^i \phi_k(\mathbf{r}). \tag{2.12}$$

Instead of finding the functional form of the Kohn-Sham orbitals, the expansion coefficients, c_k^i, have to be determined by expressing the Kohn-Sham equations (Eq. 2.8) in the basis $\phi_k(\mathbf{r})$. The resulting matrix equations have to solved in an iterative way via self-consistency.

When E_{xc} has been approximated and the starting density has been constructed, e.g. through a superposition of atomic densities, the orbitals can be generated via a self-consistent approach [Atkins and Friedman, 2004b]. When the orbitals are determined, the density can be updated via the definition of $\rho(\mathbf{r})$ (Eq. 2.7). Hence, the first step to determining the electronic energy via DFT is the approximation of the exchange-correlation functional. A widely used approximation is the local density approximation (LDA):

$$E_{xc} = \int \rho(\mathbf{r})\epsilon_{xc}[\rho(\mathbf{r})]d\mathbf{r}. \tag{2.13}$$

LDA approximates the exchange-correlation energy density per particle (Eq. 2.13), $\epsilon_{xc}[\rho(\mathbf{r})]$, of an inhomogeneous system with the one of a homogeneous electron gas (at constant density) [Atkins and Friedman, 2004b]. The correlation part of the exchange-correlation energy per electron has been determined via highly accurate quantum Monte Carlo calculations by Ceperley and Alder [Ceperley and Alder, 1981]. The exchange part is calculated analytically [Dirac, 1930; Mahan, 2013].

Although LDA is used for many problems in solid-state physics, it is not always appropriate for the description of systems in which the electron density varies a lot, such as molecular systems. In order to better account for these inhomogeneities, LDA can be augmented by gradient corrections, which results in the generalised gradient approximations (GGA). There are many variants of exchange-correlation functionals within the GGA approach. Currently, one of the popular functionals for chemical systems is the BLYP functional, which combines the gradient-enhanced exchange energy approximation of Becke with the gradient-enhanced correlation energy approximation of Lee, Yang, and Parr [Becke, 1988; Lee et al., 1988].

2.2.2 Car-Parrinello Molecular Dynamics

Car-Parrinello Molecular Dynamics (CPMD) is the method that solved the major efficiency problem of QM MD. CPMD includes the electronic structure calculation in the MD method by extending the original Lagrangian (Eq. 2.2) by introducing electronic degrees of freedom as additional classical variables with a fictitious (classical) kinetic energy [CPMD]:

$$\begin{aligned} L_{CP}(\mathbf{r}^N, \dot{\mathbf{r}}^N, \mathbf{R}^M, \dot{\mathbf{R}}^M) = \sum_{I=1}^{M} \frac{1}{2} m_I \dot{\mathbf{R}}_I^2 + \sum_{i=1}^{N} \frac{1}{2} \mu_i \langle \dot{\psi}_i \mid \dot{\psi}_i \rangle - E^{KS}(\mathbf{r}^N, \mathbf{R}^M) \\ - V_{nuc-nuc}(\mathbf{R}^N) + \sum_{i,j=1}^{N} \Lambda_{i,j}(\langle \psi_i \mid \psi_j \rangle - \delta_{i,j}), \end{aligned} \tag{2.14}$$

where μ_i is a fictitious mass for the electronic degrees of freedom and the last term ensures orthonormality of the electronic (one particle) wavefunctions ψ_i via the Lagrange multipliers $\Lambda_{i,j}$. The equations of motion for this extended Lagrangian can be obtained from the Euler-Lagrange equation in equation 2.1 in combination with the Hellmann-Feynman theorem [Atkins and Friedman, 2004c], which results in two equations of motion, one for the electronic degrees of freedom (Eq. 2.15) and one for the nuclei (Eq. 2.16) [Meijer, 2007]:

$$\mu_i \,|\ddot{\psi}_i\rangle = H^{KS} \,|\psi_i\rangle + \sum_{j=1}^{N} \Lambda_{ij} \,|\psi_j\rangle, \tag{2.15}$$

$$m_I \ddot{\mathbf{R}}_I = -\frac{\delta(E^{KS} + V_{nuc-nuc})}{\delta \mathbf{R}_I}. \tag{2.16}$$

When $\mu_i << m_I$, the extended Lagrangian (Eq. 2.14) approaches the true classical Lagrangian (Eq. 2.2), which results in exact dynamics of a real classical system. The difference in mass between the nuclei and the electronic degrees of freedom can be tuned via the fictitious mass of the electrons, μ_i, which controls how fast the electronic wavefunctions ψ_i (respectively the c_k^i coefficients in a given basis) can adapt to changes in the nuclear positions. The basis set expansion coefficients, c_k^i, are coupled to the positions of the nuclei, $\mathbf{R}_\mathrm{I}$, via the potential energy, V($\mathbf{r}$, $\mathbf{R}$). When the fictitious mass of the electronic degrees of freedom is small enough, the response of c_k^i to changes in the position of $\mathbf{R}_\mathrm{I}$ after one time step is so fast that the electronic wavefunctions remain close to the ground state. When the mass of the electrons has to be decreased also the time step decreases, which results in a longer time overall to simulate a given time window. μ_i should always be chosen in such a way that the minimum electronic frequency is significantly larger than the maximum frequency related to the nuclear motion, preventing energy transfer between the electronic and nuclear system, while using the largest μ_i possible to enable the use of a larger time step.

Besides the fictitious electronic mass, the representation of the single-particle wavefunctions influences the efficiency of the simulations. In CPMD, plane waves (PW) are used to built the basis set. PWs have several advantages as, for instance, they are unbiased because the basis set does not depend on the atomic species or position and they are consistent with periodic boundary conditions. However, PWs have one disadvantage, which is the necessity of a large number of PWs to obtain an appropriate representation of ψ_i (Eq. 2.12) [Carloni et al., 2002]. Due to the highly localised nature of core electrons, it is expensive to represent them in a PW basis set. Since the rapidly varying wavefunction of core electrons requires high energy cutoffs, the effect of core electrons are usually integrated out by using pseudopotentials and only the (slower varying) wavefunctions of the valence electrons are expanded into PWs. This is justified by the fact that valence electrons play a dominant role in the chemical behaviour of the system during reactions, while the distribution of the electrons near the nuclei do not change significantly. Using pseudopotentials, the number of PWs included in the basis set can be drastically lowered and consequently the computational cost is reduced without significant loss in accuracy.

2.3 Force-Field Based Molecular Dynamics

When the studied system contains more than a few thousands of atoms, a quantum mechanical approach cannot be applied anymore as the computational cost becomes too high. In order to reduce the expense of the calculations, one can choose to apply a molecular mechanics method or classical MD based on a parameterised potential energy surface, using experimental and theoretical data, a so-called force field. This approach describes interatomic interactions via bonded terms as well as through electrostatic interactions of effective point charges and Van der Waals interactions, described by Lennard-Jones potentials, instead of describing atoms as nuclei with explicit electrons:

$$V_{total} = V_{non-bonded} + V_{bonded}. \tag{2.17}$$

The part of the force field that is computationally the most expensive is the calculation of the non-bonded interactions. In order to decrease the computational cost of classical MD, several approximations can be applied. For example, in united-atom force fields not all atoms have to be described by explicit interaction sites because specific hydrogen atoms can also be treated in an implicit manner in order to decrease the computational cost. This unification results in non-polar hydrogens being excluded from the simulation as single particles and being included in the atomic mass of the atom originally connected to the hydrogen.

During this doctoral study the AMBER99SB force field was used to sample the thermally accessible conformations of soluble proteins and G-protein-coupled receptors [Case et al., 2015; Sorin and Pande, 2005; DePaul et al., 2010]. In this all-atom force field the non-bonded interactions are obtained by using a Lennard-Jones potential for specifying the Van der Waals interactions and a term involving fixed effective point charges to calculate the electrostatic interactions between particles (Fig. 2.1):

$$V_{LJ}(r_{ij}) = \sum_{\mathrm{i>j}} \left[\left(\frac{\mathrm{A_{ij}}}{\mathrm{r_{ij}^{12}}} \right) - \left(\frac{\mathrm{B_{ij}}}{\mathrm{r_{ij}^{6}}} \right) \right], \tag{2.18}$$

$$V_C(r_{ij}) = \sum_{i>j} \frac{q_i q_j}{\epsilon r_{ij}}. \tag{2.19}$$

In equations 2.18 and 2.19, the potentials sum over all atom pairs that experience non-bonded interactions. The Lennard-Jones potential, $V_{LJ}(\mathrm{r}_{ij})$, between atom pair $i-j$ with parameters A_{ij} and B_{ij} takes effectively into account Pauli repulsion and attractive dispersion interactions. In the representation of the Coulombic interactions, q_i and q_j are the effective charges of two particles that interact via a medium with dielectric constant ϵ [Leach, 2001].

The bonded interactions are divided into three groups: bond stretching, angle bending and dihedral terms (Fig. 2.1). The first term, bond stretching (Fig. 2.1), is describing the changes in atomic bond lengths via a harmonic potential

$$V_{bonds} = \sum_{bonds}^{Z} K_{d_Z}(d_Z - d_{Z0})^2, \tag{2.20}$$

with d being the bond length that is compared to the reference bond length d_0. In an analogous way, the bond angle potential (Fig. 2.1) of the AMBER force field is described by

$$V_{angles} = \sum_{angles}^{L} K_{\theta_L}(\theta_L - \theta_{L0})^2. \tag{2.21}$$

Through this potential, the energy increase due to a deviation of the angle, θ, with respect to the reference value θ_0 is obtained. The last term to complete the bonded potential is the dihedral potential (Fig. 2.1), which is defined by

$$V_{dihedrals} = \sum_{dihedrals}^{B} \sum_{n} \frac{V_n}{2}[1 + \cos(n\phi_B - \gamma_B)], \tag{2.22}$$

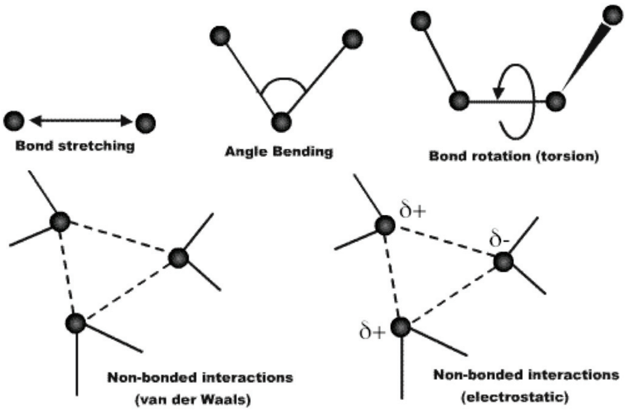


Figure 2.1: Schematic image of the bonded and non-bonded interactions that are being taken into account during classical MD. The non-bonded interactions are defined by the electrostatic and Van der Waals interactions. The bonded terms can be described by harmonic bond stretching concerning the distance between a pair of atoms i and j, d_Z, and angle bending concerning the angle, θ_L, between three particles i, j, and k. The last term in the bonded segment includes the torsion between particles, which is calculated via the angle ϕ between the two planes of consecutive triples of atoms i, j, k and j, k, l. Taken from Leach [2001].

with V_n being a barrier for dihedral transitions, n denotes the periodicity and γ represents the phase shift of the torsional angle ϕ [Cornell et al., 1995; Weiner et al., 1984].

2.4 QM/MM Molecular Dynamics

The idea of combining QM and MM MD to enhance both accuracy and efficiency of large system simulations, has been a challenge for several decades and is still a topic of development [Warshel and Levitt, 1976; Singh and Kollman, 1986; Field et al., 1990; Laio et al., 2002a; Bulo et al., 2009]. The QM/MM approach is important for simulations of biomolecules that are too large for QM MD but for which classical MD is not adequate to simulate the chemical reactions that can occur in an active site, for example. Hence, a combination of QM and classical MD is applied to study these reactive proteins (Fig. 2.2 a). The Hamiltonian of the system is

$$H^{total} = H^{QM} + H^{MM} + H^{QM/MM}. \tag{2.23}$$

In equation 2.23, H^{QM} is the quantum mechanical Hamiltonian of the QM part (e.g. the active site), H^{MM} the molecular mechanics Hamiltonian for the MM part of the protein and the solvent, and $H^{QM/MM}$ describes the interaction between the QM and the MM part. Consistent with the functional form of the force field, this term is divided into non-bonded and bonded interactions. The bonded terms are included when the QM/MM interface separates atoms from one another that are connected via covalent bonds (Fig. 2.2 b). After the cut between the QM part and the MM part, the QM region can become electronically unsaturated if covalent bonds exist between the QM and the MM region. One approach that can be used to saturate

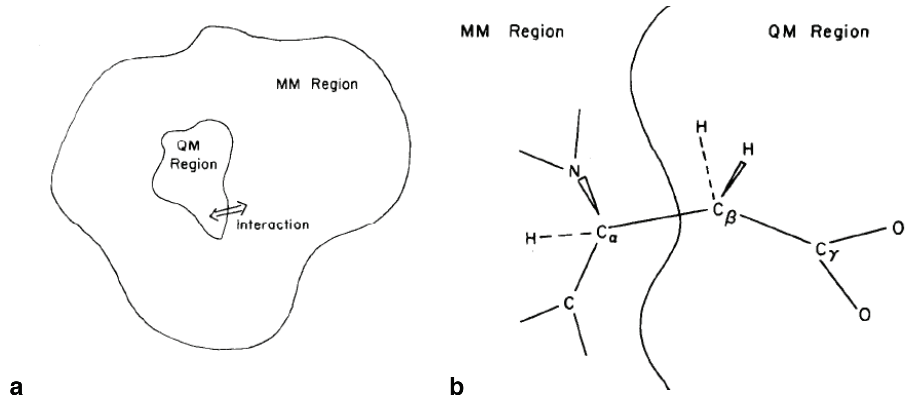


Figure 2.2: (a) Illustration of the devision between the QM and MM region in the system. Taken from Field *et al.* [Field et al., 1990]. (b) Representation of the interface between the QM and MM region in which two atoms are connected to one another, one located in the MM part and the other in the QM part. Taken from Field *et al.* [Field et al., 1990].

the valence orbitals of the QM part and to prevent an unphysical description of the atomic orbitals at the border of the QM region, is hydrogen capping in which link atoms (hydrogen atoms) are added to the QM region [Laio et al., 2002a]. The MM region of the system treats the last QM atom as part of the MM region by using force field calculations to obtain its potential energy and the link atom is not detected by the MM system. In addition, the angular and dihedral terms of the boundary Hamiltonian remain to be described by the MM force field. Another approach to solve the problem for QM/MM interfaces cutting through covalent bonds is to apply parameterised monovalent pseudopotentials to atoms at the QM border that are part of a covalent bond with the MM region [von Lilienfeld et al., 2005].

The non-bonded interactions between the QM part and the MM region are defined by electrostatic and steric effects. The Van der Waals interaction term of the MM region is used to describe the steric non-bonded interactions, which are a result of the Pauli repulsion and dispersion interactions. Together with the steric effect, also the bonded interactions are generated by the molecular mechanics part of the system. However, according to the applied type of electrostatic coupling, the electrostatic effect is included in the quantum mechanical system as an addition to the external field [Carloni et al., 2002]. These electrostatic interactions between the classical point charges and the quantum mechanical electron density have a significant impact on the QM system. The general non-bonded interaction Hamiltonian is

$$H^{QM/MM}_{non-bonded} = V^{QM/MM}_{C} + V^{QM/MM}_{vdw}, \tag{2.24}$$

which includes the Van der Waals and Coulomb interaction terms and can be defined as:

$$H^{QM/MM}_{non-bonded} = \sum_{i \in MM} q_i \int d^3 r \rho(\mathbf{r}) v_i(|\mathbf{r} - \mathbf{r}_i|) + \sum_{i \in MM,\ j \in QM} v_{vdw}(\mathbf{r}_{ij}). \tag{2.25}$$

For this Hamiltonian i represents the MM atom for which the charge is q_i and the position is $\mathbf{r}_i$ [Laio et al., 2002a]. ρ in equation 2.25 is the total charge density (of the nuclei and electrons) of the quantum mechanical system, v_{vdw} is the Van der Waals interaction between atom i (MM) and atom j (QM). v_i is a function for the Coulomb interaction modified at short-range

$$v_i(r) = \frac{r^n_{ci} - r^n}{r^{n+1}_{ci} - r^{n+1}}, \tag{2.26}$$

with r_{ci} representing the covalent radius of the atom i, $r = |\mathbf{r} - \mathbf{r}_i|$ and n=4.

This representation of $H^{QM/MM}_{non-bonded}$ in equation 2.25 provides a solution for the short-range interaction problem, which occurs between the QM and MM region. This short-range interaction problem or electron spill-out problem is present because positively charged atoms in the

MM part of the system are able to overattract the electron density of the QM atoms, which is a result of the absence of the Pauli repulsion of the MM atom as it does not contain electron density. This can cause an overpolarisation of the electron density as the positively-charged MM atom only contains an exclusively attractive potential and provides an incomplete representation of the interactions [Laio et al., 2002a]. This form of $H^{QM/MM}_{non-bonded}$ used in the CPMD package (Eq. 2.25) is able to prevent the electron spill-out problem.

However, calculating the electrostatic interactions via equation 2.25, which requires summing over all charges of the MM region interacting with the electron density at every real-space grid point, would become computationally too expensive when the simulated system includes 100.000 or 200.000 atoms, a common size for protein systems and of the order of 100^3 grid points. Therefore, in large systems the Coulomb interaction is only calculated as in equation 2.25 [Laio et al., 2002b,a] for MM atoms within a particular radius from each QM atom, the radius$_{NN}$ (Fig. 2.3).

For MM atoms further away from the QM region that are not included in the NN set but are part of the ESP region, radius$_{NN} \leq$ radius $<$ radius$_{ESP}$, the Coulomb interaction term is altered to reduce the computational cost. To this end, the electrostatic potential of the NN atoms (Eq. 2.25) can be used as a target to fit effective point charges on the QM atoms that reproduce the electrostatic quantum potential that was calculated for the QM region with the NN atoms as probe sites [Laio et al., 2002b]. These charges are called Dynamically Restrained Electrostatic Potential (D-RESP) derived charges as the charges are obtained during every step of the MD trajectory and the fit is restrained through a weighting function that depends on the Hirshfeld

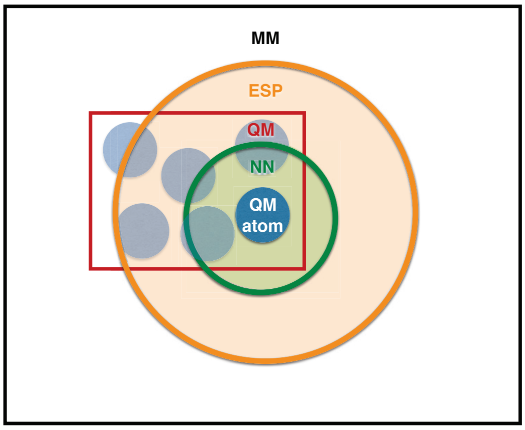


Figure 2.3: Schematic representation of the regions of MM atoms (NN in green and ESP in orange) around an example QM atom that are used to calculate $H^{QM/MM}_{non-bonded}$. The red rectangle represents the QM region while the blue filled circles depict QM atoms in the QM region.

charges to prevent large fluctuations in the charges due to conformational changes [Laio et al., 2002b]. The Coulomb interaction term for the ESP region can thus be defined as:

$$V_C^{ESP} = \sum_{i\in ESP, j\in QM} \frac{q_i q_j^{D-RESP}}{r_{ij}}, \tag{2.27}$$

where q_i is the classical point charge on an MM atom i in the ESP region, q_j^{D-RESP} is the fitted point charge on a QM atom j and r_{ij} is the distance between $\mathbf{r}_i$ and $\mathbf{r}_j$. When the distance between the MM and every QM atom becomes sufficiently large –larger than radius$_{ESP}$– the Coulomb interaction of the MM atoms with the QM region can be expanded into a multipolar expansion [Laio et al., 2002a]:

$$V_C^{MM} = \sum_{i\in MM} q_i \Big[C \frac{1}{|r_i - \bar{r}|} + \sum_{\alpha} D^{\alpha} \frac{(r_i^{\alpha} - \bar{r}^{\alpha})}{|r_i - \bar{r}|^3} + \tag{2.28}$$

$$\frac{1}{2} \sum_{\alpha\beta} Q^{\alpha\beta} \frac{(r_i^{\alpha} - \bar{r}^{\alpha})(r_i^{\beta} - \bar{r}^{\beta})}{|r_i - \bar{r}|^5} + \mathcal{O}\Big(\frac{R_q}{|r_i - \bar{r}|}\Big)\Big].$$

In equation 2.28 $\bar{r}$ is the origin of the multipolar expansion defined as the geometrical center of the QM region. C, D^{α} and $Q^{\alpha\beta}$ are the total charge, the dipole, and the quadrupole of the quantum charge distribution, ρ, of which the detailed description can be found in Laio et al. [2002a]. The charge density of the QM region is located within R_q, the radius of a sphere around the geometrical centre of the QM region, $\bar{r}$ [Laio et al., 2002a]. Hence, the complete non-bonded interaction Hamiltonian that is used in the QM/MM interface of the CPMD package to describe the interactions between the MM and the QM atoms in large systems is:

$$H_{non-bonded}^{QM/MM} = V_C^{NN} + V_C^{ESP} + V_C^{MM} + \sum_{\substack{i\in NN,ESP,MM \\ j\in QM}} v_{vdw}(\mathbf{r}_{ij}). \tag{2.29}$$

2.5 Force Matching

The most commonly used force fields in biomolecular simulations have been parametrised using experimental data sets or quantum mechanical calculations. However, molecular moieties that are non-standard components of biological molecules, unlike amino acid residues, for example, require individual parameterisation of the moiety as accurate parameters are usually not present in the parameter set of the force field. AMBER released a general AMBER force field

(GAFF) [Wang et al., 2004], which can be used for pharmaceutical compounds. However, GAFF does not provide customised parameters for molecules, but assigns atom types that resemble best the atoms in the molecule. Generating customised parameters for a molecule is time consuming and can result in non-transferable parameters when the missing parameters of the molecule highly depend on the environment. This is especially the case when the parameters are determined from electronic structure calculations of small molecules in a continuum solvent or in the gas phase [Bayly et al., 1993].

The QM/MM force-matching approach implemented in the CPMD package by Rothlisberger *et al.* [Doemer et al., 2013; Maurer et al., 2007] is a technique that is able to parametrise a molecule in its native environment, such as a protein, by using a QM/MM MD reference trajectory. Assuming that the force field used on the MM region adequately represents this part of the system, the risk of non transferability in the automatised force-matching procedure is significantly lower as the molecule's environment is taken into account during the parameter calculations. The calculations are performed on a reference trajectory obtained via QM/MM MD at finite temperature with the QM region including the molecular group to be parameterised. After generating the QM/MM MD reference trajectory, three steps have to be performed to obtain the new parameters for the molecular moiety in the QM region. The first step involves the storing of all the forces on the atoms in the QM region as well as the electrostatic potential and field on the MM atoms that are within a certain radius of the QM atoms, the NN region (see Section 2.4). In the third step, atomic point charges q_i are calculated for the QM atoms to reproduce the electrostatic potential and field obtained in step two. The final step covers the force matching of the remaining parameters of the molecule in the QM region, apart from the Van der Waals interactions that are taken from the original force field.

The force matching routine can be split into two steps. First the non-bonded contributions to the atomic forces, computed via the obtained atomic point charges q_i^{RESP} from the QM region and Lennard-Jones parameters for the QM region, taken from the MM force field, are subtracted from the total QM reference forces. In the second step the parameters for bond stretching, angle bending and dihedral torsion (see Section 2.3) are modified in such a way as to reproduce the remaining force components.

2.5.1 Step 1 and 2: Reference Trajectory and Forces

The QM/MM MD trajectory from which frames are used as reference configurations to calculate the forces, the electrostatic potential and field on the nearby MM atoms is generated with the CPMD package using QM/MM Car-Parrinello MD or Born-Oppenheimer MD. The time scale that can be reached via this approach lies in the range of picoseconds which is usually too short to observe spontaneous dihedral transitions. Therefore, the force field only represents a particular configuration of the molecule and additional torsional sampling via enhanced sampling techniques has to be performed in order to fully fit these parameters.

When Car-Parrinello MD is employed to propagate the system, the system is close to the ground state, however, accurate ground-state nuclear forces are needed to perform force matching. Consequently, the obtained Car-Parrinello MD configurations are used to compute BO reference forces for each configuration with tight convergence criteria. Subsequently, these BO reference forces are used to fit a force-field parameter set for the atoms in the QM region.

2.5.2 Step 3: Fit of Atomic Point Charges of QM Region

After obtaining the reference forces of the QM region in the QM/MM MD trajectory as well as the electrostatic potential and field to which the NN atoms are exposed to (see Section 2.4), atomic point charges of the QM atoms are fitted to reproduce the electrostatic potential and the electric field on the NN atoms that is a result of the QM charge density. This is a modified scheme from the D-RESP procedure (see Section 2.4) [Laio et al., 2002b, 2004], used in calculating the non-bonded interactions during QM/MM MD, because besides the electrostatic potential also the electric field on the NN atoms has to be reproduced by the fit of the atomic point charges. Another important modification to the original D-RESP procedure in the force matching scheme is that the charges are not calculated per configuration but a single set of atomic point charges on the QM atoms is fitted that performs best on average over all QM/MM MD reference configurations included in the fit. The incorporation of several QM/MM MD configurations in the charge fit leads to more transferable charge sets compared to charges obtained via a single configuration [Reynolds et al., 1992]. As a result, a weaker restraint to the Hirshfeld charges can be used compared to atomic point charge calculations on one configuration [Maurer et al., 2007]. A restraint to the Hirshfeld charges is necessary to avoid chemically unreasonable and conformation dependent solutions of the point charge fit [Laio et al., 2002b].The fit of the atomic point charges q_i is obtained by minimising the penalty function [Doemer et al., 2013; Maurer et al., 2007], χ^2:

$$\chi^2(\{q_j\}) = \sum_{l=1}^{L} \Big[\sum_{i \in NN_i} \Big(w^V (V_{il}^{MM} - V_{il}^{\rho})^2 + \sum_{i \in NN_l} w^F \, || \, \mathbf{E}_{il}^{MM} - \mathbf{E}_{il}^{\rho} \, ||^2 \Big) + \sum_{j \in QM} w^Q (q_j - q_{jl}^H)^2 \Big] + w^{tot} (Q^{tot} - \sum_{j \in QM} q_j)^2, \tag{2.30}$$

with l being the index running over all reference configurations L, j representing a classical atom from the NN_l set which are explicitly coupled to the quantum charge density of configuration l via equation 2.30 and i including all QM atoms in the system. The different weighting factors, w, which can be tuned in the force-matching procedure to balance the different contributions [Doemer et al., 2013], stand for weighting factors of the electrostatic potential (w^V), the electric field (w^F), the Hirshfeld charges (w^Q) and the total charge of the QM subsystem (w^{tot}). V_{il}^{ρ} an $\mathbf{E}_{il}^{\rho}$ are the electrostatic potential and field, respectively, on the classical atom i in configuration l due to the presence of the QM system. V_{il}^{MM} and $\mathbf{E}_{il}^{MM}$ are the electrostatic potential and field resulting from the classical point charges $\{q_j\}$. The third restraint on the charges keeps the charges q_j close to the Hirshfeld charges q_{jl}^{H}, while the final restraint ensures that the total charge of the QM region adds up to the correct value Q^{tot}.

2.5.3 Step 4: Fit of Bonded Interactions of the QM Region

After fitting a set of atomic point charges $\{q_j\}$ on the QM atoms, the total non-bonded forces, $\mathbf{F}_{jl}^{MM_{nb}}$, of QM atom j in configuration l are calculated by including the fitted charge as well as the MM Van-der-Waals interactions. These newly obtained non-bonded forces of the QM region are subtracted from the QM/MM reference forces, $\mathbf{F}_{jl}^{QM}$, leading to the part of the forces due to the bonded interactions in the QM region. The bonded interactions described by a set of parameters $\{x_n^b\}$ [Maurer et al., 2007] are fitted according to the penalty function, σ^2:

$$\sigma^2(\{x_n^b\}) = \sum_{l=1}^{L} \sum_{j \in QM} \| \mathbf{F}_{jl}^{MM_b} - \left(\mathbf{F}_{jl}^{QM} - \mathbf{F}_{jl}^{MM_{nb}} \right) \|^2 .$$

The newly obtained bonded parameters together with the new atomic point charges and the MM Van-der-Waals interactions from the original force field will be used to compose a new parameter set for the target molecule, which can be used for long time propagation via classical molecular dynamics simulations of the specific system.

3 How Rhodopsin Tunes the Equilibrium between Protonated and Deprotonated Forms of the Retinal Chromophore

Siri Camee van Keulen[1], Alicia Solano[2], Ursula Röthlisberger

3.1 Summary

Rhodopsin is a photoactive G-protein-coupled receptor (GPCR) that converts dim light into a signal for the brain, leading to eyesight. Full activation of this GPCR is achieved after passing through several steps of the protein's photoactivation pathway. Key events in the activation of rhodopsin are the *cis-trans* photoisomerisation of the retinal moiety followed by conformational rearrangements and deprotonation of the protonated Schiff base (PSB) which lead to full activation in the meta II state. Deprotonation of PSB is crucial for achieving full activation of rhodopsin, however, the mechanism that takes place to reach the deprotonated state is not well understood. In this study the molecular factors that enable PSB deprotonation at body temperature in the later stages of the photoactivation process have been investigated in order to obtain a better understanding of the difference between the deprotonatable state of the chromophore and the inactive conformation.

---

[1] Van Keulen performed the discussed classical MD and QM/MM MD simulations, the ZINDO/S calculations and the pKa calculations of the dark state and the early intermediates of rhodopsin.

[2] Solano performed the discussed correlation and causality calculations of the early intermediates in comparison to the dark state of rhodopsin.

The deprotonation mechanism has been investigated by applying QM/MM simulations in combination with thermodynamic integration. The study shows that, compared to the inactive 11-*cis*-retinal case, *trans*-retinal rhodopsin is able to undergo deprotonation of PSB due to a change in the conformation of the retinal and a consequent alteration in the hydrogen-bond (HB) network in which PSB and Glu113 are embedded. Besides the retinal moiety and Glu113, also two water molecules as well as Thr94 and Gly90 that are related to congenital night blindness are part of this HB network.

3.2 Introduction

In vertebrates' eyes, rod cells are activated when only dim light is available. The protein that is able to convert these photons into a signal for the brain is rhodopsin. Rhodopsin is a protein with seven trans-membrane domains (TMs) and is part of class A in the G-protein-coupled receptor (GPCR) family. For several decades rhodopsin has been intensely studied as a prototype system to understand the activation process of GPCRs. Besides being viewed as an example protein for other GPCRs in class A, rhodopsin bears also other properties of interest, such as low basal activity and a high yield of around 65% for the conversion of its inactive 11-*cis*-retinal state (Fig. 3.1) to its active *trans* configuration [Kim et al., 2001]. However, these features can be perturbed via mutations in the active site of rhodopsin, such as Thr94Ile and Gly90Asp, that cause congenital stationary night blindness (CSNB) [Singhal et al., 2016, 2013].

Besides mutations, rhodopsin is also affected by changes in temperature that lead to different activation pathways at low temperatures with respect to room temperature. For example, depending on the temperature of the system, the deprotonation of the protonated Schiff base (PSB) is shifted to an earlier (metarhodopsin I) or later step (metarhodopsin II) in the photoactivation pathway (Fig. 3.1b) [Lewis et al., 2004; Sandberg et al., 2014; Thorgeirsson et al., 1993]. Besides the change in PSB deprotonation, the blue-shifted intermediate (BSI) is absent at low temperatures while at room temperature BSI is in equilibrium with the batho state (Fig. 3.1b)[Hug et al., 1990; Lewis and Kliger, 1992]. Several structures of intermediates of the photoactivation pathway, dark, batho, lumi and meta II [Nakamichi and Okada, 2006a,b; Okada et al., 2004; Li et al., 2004; Deupi et al., 2012], have been obtained at low temperatures via X-ray crystallography (160 Kelvin) and show the relaxation after *cis*-*trans* photoisomerisation. However, in view of the fact that in nature rhodopsin functions at body temperature and the observed temperature dependence of the activation path, actual early intermediates under physiological conditions could differ from the configurations found via X-ray crystallography.

Rhodopsin's activation pathway starts with inactive, dark state, rhodopsin in which Glu113 forms a salt bridge with the protonated Schiff base. However, after photon exposure, relaxation of the retinal moiety and deprotonation of the chromophore, the salt bridge is broken and remains broken in the long-lived later intermediates such as meta II (Fig. 3.1b) [Cohen et al., 1992]. During the deprotonation step, a proton is transferred from the protonated Schiff base, formed by the retinal and the covalently bound Lys296, to Glu113, the counter ion [Jaeger et al., 1994; Sakmar et al., 1989]. A parameter that plays an important role in the deprotonation mechanism is the pKa of retinal's protonated Schiff base. PSB's pKa, which is estimated to be greater than 15 in dark state bovine rhodopsin [Ebrey, 2000; Steinberg et al., 1993], can be decreased to 6 by mutating Glu113 to a glutamine residue [Sakmar et al., 1991], which is close to PSB's pKa value of 7.5 ± 0.1 in solution [Gat and Sheves, 1993]. The effect of the Glu113Gln mutation can thus trigger the deprotonation of the Schiff base and shows that the change from a charged residue to a polar one near the PSB can affect the chromophore's protonation state [Zhukovsky et al., 1992; Gat and Sheves, 1993; Ebrey, 2000].

As temperature dependence plays a significant role, the available X-ray structures of the early

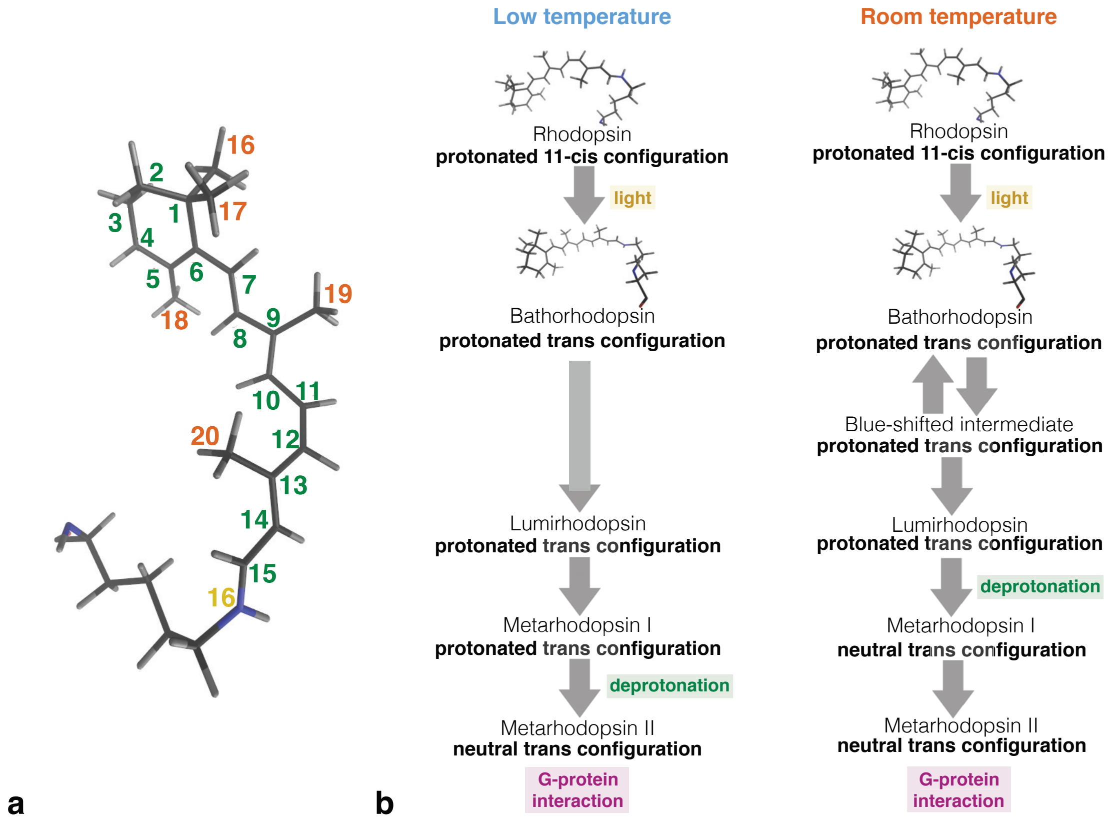


Figure 3.1: Rhodopsin's photoactivation pathway. (a) Structure of the retinal moiety in the 11-*cis* configuration covalently linked to the protonated Schiff base (Lys296). The number of the methyl groups are depicted in orange, while the number of the nitrogen is shown in yellow. The numbers of the residual carbons of the retinal moiety are represented in green. (b) Schematic representation of the photoactivation pathway of rhodopsin at low (between 10 Kelvin to 273 Kelvin) and room temperature [Sandberg et al., 2014; Lewis et al., 2004; Hug et al., 1990].

intermediates might not be appropriate structures to start a molecular dynamics study with as they were obtained at low temperatures [Nakamichi and Okada, 2006a,b]. A feature of the active site that can be affected by a temperature increase, for example, is the sampling of different type of configurations during the relaxation of the *trans* retinal configuration [Hug et al., 1990; Lewis and Kliger, 1992]. The positioning and the number of water molecules is a second important factor that can be altered by the temperature of the system as well as by the crystallisation process.

In order to investigate PSB deprotonation at body temperature, a dark state X-ray structure of rhodopsin was used to build an initial homodimer structure to minimise temperature and crystallisation effects on the activation mechanism of the protein. A homodimeric form of rhodopsin was employed in the classical MD simulations in order to mimic the native environment of rhodopsin in the retina [Fotiadis et al., 2006]. Besides in retina, homodimerisation has also been observed for rhodopsin:G-protein complexes for which a stoichiometry of 2:1 could be determined [Jastrzebska et al., 2011, 2013]. After classical MD simulations, QM/MM MD simulations were used to follow the relaxation process during the early intermediates (batho, BSI, lumi) as well as to investigate the deprotonation reaction in the lumi state via thermodynamic integration (TI). The validity of the configurations of the early intermediates was checked through the comparison of their calculated vertical excitation energies with the corresponding experimental absorption maxima. This comparison showed that the found configurations were in good agreement with experiment.

To anticipate our results, the QM/MM MD at body temperature shows that in the early intermediates the relaxation of the retinal conformation mainly takes place close to PSB where space in the active site is present for methyl relocation. The comparison of PSB deprotonation in the dark and the lumi state suggests that besides the orientation of the retinal moiety also the hydrogen-bond network around the PSB is important for the feasibility of the reaction. At body temperature, water molecules in the lumi state appear to impact this HB network severely by changing the salt bridge between PSB and Glu113 to a bridged water form thereby decreasing Glu113's hydrogen-bond accepting interactions from four to three. After the deprotonation reaction, Glu113's hydrogen bonds are even decreased to two, of which one Glu113 remains the hydrogen-bond acceptor and in the second one Glu113 is the hydrogen-bond donor. This investigation suggests that the retinal orientation within the active-site pocket as well as the rearrangement of the HB network that takes place during the relaxation of the retinal moiety, are important to favour PSB deprotonation.

3.3 Methods

3.3.1 Classical Molecular Dynamics

The periodically repeated simulation box has a size of approximately 120 X 115 X 126 Å^3, which includes around 40 000 water molecules, 400 lipids and four Na$^+$ ions to neutralise the rhodopsin dimer system (Fig. A.1). The lipids are 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamines (POPEs) for which the force field parameters of Jämbeck *et al.* were used [Jämbeck and Lyubartsev, 2012]. The rhodopsin homodimer system was simulated at 310 K and at a pressure of 1 bar using a Nosé-Hoover thermostat and a semiisotropic Parrinello-Rahman barostat. The initial structure of the rhodopsin monomers was taken from the protein data bank (PDB) [Berman et al., 2000], PDB code 1U19 [Okada et al., 2004], while the missing residues 214-246 were taken from PDB code 1GZM [Li et al., 2004]. The intradimeric TM4/TM5 interface was used for the homodimer, extracted from PDB code 1N3M [Fotiadis et al., 2006]. Default protonation states were used for the residues in the monomers, except for Asp83 and Glu122 that were protonated [Fahmy et al., 1993]. A disulfide bond was formed between Cys110 and Cys187. Two palmitate groups, for which force field parameters were taken from literature [Röhrig et al., 2002; Dickson et al., 2012] are bound to each monomer via Cys322 and Cys323. The force field parameters for the retinal and Lys296 used in this study were obtained by Saam *et al.* [Saam et al., 2002].

The force fields used for the protein dimer and the water molecules are AMBER99SB [Case et al., 2015] and TIP3P [Jorgensen et al., 1983], respectively. All MD runs were perofrmed using GROMACS 4.5.5 [Hess et al., 2008]. Electrostatic interactions were calculated with the Ewald's particle mesh method with a real space cutoff of 12 Å. Bonds involving hydrogen atoms were constrained using the LINCS algorithm. The integration time step was set to 2 fs.

The production run of the rhodopsin dimer system in the dark state was performed for 700 ns after 10 ns of equilibration. Subsequently, the classical MD simulation of the dark state dimer was used to initiate *cis*-*trans* isomerisation of a single retinal. The use of classical MD simulations to perform the restrained conversion from *cis* to *trans* was previously validated by Röhrig *et al.* [Röhrig et al., 2004, 2005] and shows that excited state calculations lead to a very similar configuration of the isomerised retinal as restrained classical MD simulations. The *cis*-*trans* isomerisation was performed in one monomer by putting restraints on the C10-C11-C12-C13 and C11-C12-C13-C14 dihedrals to secure the desired *trans* configuration (Fig. 3.1a). After the isomerisation has taken place in around 200 fs, the restraints were released and snapshots were taken that were used as initial structures for the QM/MM simulations.

3.3.2 Quantum Mechanics/Molecular Mechanics Molecular Dynamics

The QM part of the QM/MM MD simulations included the retinal moiety, Lys296, Glu113 and two nearby water molecules (Fig. A.2). The CPMD 3.15 package [CPMD] was employed to perform the QM/MM MD simulations with Car-Parrinello molecular dynamics (CPMD) using the BLYP functional and a plane wave energy cutoff for the wavefunction of 70 Ry. Martins-Troullier pseudopotentials were used for the QM atoms. In order to cut the QM residues from the rest of the protein, monovalent pseudopotentials [von Lilienfeld et al., 2005] were employed. QM/MM simulations were performed at an average temperature of 310 K using a Nosé-Hoover thermostat. A time step of 0.12 fs and a fictitious electronic mass of the CP Lagrangian of 700 au were used. The AMBER force field was used to account for interactions among the atoms of the MM region. In case of the simulations performed for thermodynamic integration, the QM region was extended, compared to the QM/MM MD for the spectra calculations, by the incorporation of the backbone of Glu113, Thr94, and the backbone of Cys187 (Fig. A.2). These residues were added to the QM region as Thr94 and Cys187 are hydrogen bonded to Glu113. The backbone of Glu113 was included because it interacts with the water molecules in the QM region (Fig. 3.2c).

In the TI simulations, the constrained distance in the dark state was defined as the distance between the proton of the protonated Schiff base and the oxygen of Glu113 that form a salt bridge with the proton (Fig. 3.2a). In the lumi state the constrained distance was described by the distance between a hydrogen of a water molecule that bridges the PSB and Glu113 and the oxygen of Glu113 that forms a hydrogen bond with the bridging water molecule (Fig. 3.2c). The interval in the constraint distance between each window was 0.1 Å except for three windows, between 1 Å and 1.15 Å for the dark and lumi state of 1 Å, 1.05 Å and 1.15 Å, in which an interval of 0.05 Å was applied. Each constraint window for both dark and lumi thermodynamic integration procedures was sampled for a minimum length of 2.4 ps, unless stated otherwise in the supplementary information.

3.3.3 pKa Calculations

pKa values for Glu113, Asp83 and Glu122 were calculated via the H++ webserver [Gordon et al., 2005; Myers et al., 2006; Anandakrishnan et al., 2012]. The input file for the pKa calculations contained a dark, lumi or pre-meta-I configuration of the rhodopsin monomer that was isomerised and included water molecules within a 7 Å radius around the retinal moiety of the rhodopsin monomer. All aspartates and glutamates were not protonated. The retinal moiety was not included in the pKa calculations as it could not be taken into account by the H++ server. In the performed calculations, the salinity was kept at 0.15 M and the external dielectric was set to 80. Values for the interior dielectric were varied from 4, 6, 8, 10 to 20.

3.3.4 Spectra Calculations

The ZINDO/S method implemented in Gaussian 09 [Frisch et al.] was used to calculate the vertical excitation energies for rhodopsin, bathorhodopsin, BSI, lumirhodopsin and pre-metarhodopsin-I configurations. Previously obtained vertical excitation energies via ZINDO/S of retinal and related polyenals have been shown to be in line with experiments [López et al., 2006]. The four lowest roots were taken into account together with their corresponding oscillator strengths in order to obtain the absorption spectra. The QM cluster model that was used to calculate the vertical excitation energies included the retinal moiety and residues within a distance of 7 Å from the retinal group. The sufficiency of the use of a 5 Å radius for the rhodopsin system has been previously determined by Campomanes *et al.* [Campomanes et al., 2014]. In addition, the backbone of each residue was removed in the model and the position of the Cα carbons was substituted for a hydrogen atom to saturate the system.

3.3.5 Correlation Feature Selection and Causality Inference Analysis

A two-step protocol was employed to identify the set of electronic and geometrical features responsible for the shifts in the batho, lumi and pre-meta I absorption spectra with respect to the dark conformation. The initial set of features was composed of the bond length alternation (BLA) [Gieseking et al., 2015] as well as all intramolecular and intermolecular distances, angles and dihedrals in and around the retinal moiety with a distance cutoff of 7 Å. The BLA is defined as the difference between the sum of all the single-bond lengths minus the sum of all the double-bond lengths along the chromophore's conjugated polyene chain, from C5 to the N atom of the protonated Schiff base (Fig. 3.1a). Within the selected region all atoms were taken into account except for the hydrogen atoms and the protein backbone. For each photointermediate, the 100 snapshots used in the ZINDO/S calculations were included and resulted in an initial set of over $48 * 10^6$ features. Since the number of features in the initial set is computationally unfeasible, the dimensionality of the initial set of descriptors had to be reduced to a smaller subset of features, which is still indispensable and sufficient to describe the target parameter, the spectral shift. The spectral shift is defined as the maximum of the calculated vertical absorption spectra of the excitation energy of the intermediate in each snapshot minus the one of the dark state (2.39 eV) (Table 3.2). Once the dimensionality problem is sufficiently reduced, causality can be inferred between the features in the reduced set and the spectral shift.

Hence, the first step towards performing the correlation and causality analysis includes reducing the number of features by removing features that are irrelevant, redundant and noisy with respect to the target property, the spectral shift. Two feature selection algorithms implemented in the WEKA 3.8 machine learning package [Witten and Frank, 2005] were employed to perform this task. Previously, WEKA has already been used successfully for an analysis of the spectral features of rhodopsin [Campomanes et al., 2014], taking into account all intermediates at once, up to the lumi state. The Regressional ReliefF (RReliefF) attribute evaluator algorithm [Robnik-Šikonja and Kononenko, 1997] together with a ranker search method was used first to reduce the dimensionality of the initial set of features to a reduced set of features. Subsequently, the Correlation Based Feature Selection (CBFS) algorithm [Hamilton, 1998], using the BestFirst search method, was employed to perform a correlation analysis on the reduced feature set to select features that are highly correlated with the spectral shift. CBFS is a filter algorithm that evaluates the heuristic merit of a subset of features considering the correlation to the target, and the correlation within features. Therefore, the selected subset, showing the highest merit, contains highly correlated features with the spectral shift, which are minimally correlated with each other. The heuristic merit is defined as:

$$M_s = \frac{k < r_{ft} >}{\sqrt{k + k(k-1) < r_{ff} >}}, \tag{3.1}$$

where M_s is the heuristic merit of a subset s, containing k features, $< r_{ft} >$ is the mean feature-target correlation, and $< r_{ff} >$ is the mean feature-feature intercorrelation. Hence, the merit numerator provides a measure of the predictive ability of the subset of features to the target, whereas the denominator gives an indication of the feature's redundancy.

After applying the CBFS algorithm on the reduced set of features, the PC-LiNGAM algorithm [Hoyer et al., 2012], implemented in the TETRAD V causal inference package [Glymour and Scheines, 1986], was used to deduce a causal relationship between the spectral shift and the features included in the set provided by the results of the CBFS algorithm. By combining two different approaches, the PC (Peter-Clark) algorithm [Spirtes et al., 2000] and the LiNGAM (Linear Non-Gaussian Acyclic Model) algorithm [Shimizu et al., 2006], the PC-LiNGAM method accepts an arbitrary distribution, meaning that Gaussian and non-Gaussian data sets can be used to obtain the structural equation model (SEM). The SEM represents the causal structure of the model in the form of linear equations and is graphically modelled by a direct acyclic graph (DAG) [Lacerda et al., 2012]. The inference of this approach is based on conditional independences between the variables. Conditional independences between variables were estimated via the Fisher's z-transform, using a significance level, α, of 0.001. Moreover, the Markov blanket of the spectral shift was built and the values for the edge coefficients in the SEM were estimated using a regression optimiser after standardisation (zero mean and unit variance) of the data set variables.

3.3.6 Structure Superpositions and Images

Multiprot [Shatsky et al., 2004] and VMD [Humphrey et al., 1996] were used to align protein structures. Images were prepared with VMD [Humphrey et al., 1996].

3.4 Results and Discussion

3.4.1 Possible Deprotonation Pathway for Nitrogen of PSB in Rhodopsin

The experimental study by Sandberg *et al.* [Sandberg et al., 2014] shows that at room temperature the deprotonation of PSB already takes place between the lumi and meta I state, instead of taking place between the meta I and meta II state, which is the case at low temperatures. Therefore, a QM/MM equilibrated lumi structure at body temperature was used to follow the deprotonation reaction via thermodynamic integration. For the sake of comparison, analogous calculations were performed for the dark state. For both configurations, the distance between Glu113's O1, which is part of the salt bridge with PSB in the dark state (Fig. 3.2), and the hydrogen to which it is hydrogen bonded was used as reaction coordinate in the TI simulations. Hence, in the lumi state the distance between H(WAT1) and O1(Glu113) was used as a constraint, while for the dark state O1(Glu113) and the hydrogen on the nitrogen of the PSB, H(PSB), was used (Fig. 3.2 and Fig. 3.3a). In the TI simulations of the lumi state, the protonation of Glu113 via the proton of WAT1 results in the spontaneous deprotonation of retinal's PSB and reprotonation of WAT1 (Fig. 3.3 and Fig. A.10, A.11), which suggests that the used reaction coordinate is a valid one.

Figure 3.4 shows that a significant difference can be found in the free energy barrier of the two states. While the lumi state reaches a plateau after deprotonation and its conformation is stable in the unconstrained simulation after TI (Fig. A.7), the dark state does not reach a minimum after deprotonation and the Schiff base is spontaneously reprotonated after the constraint is released (Fig. 3.4 and Fig. A.8, A.12, A.13). When the conformational rearrangements upon deprotonation are compared between the dark and lumi state TI simulations, it becomes apparent that in the dark state, the three hydrogen bonds, present between O1(Glu113) and the active-site environment (Fig. 3.2a), do not change during the TI simulations or in the

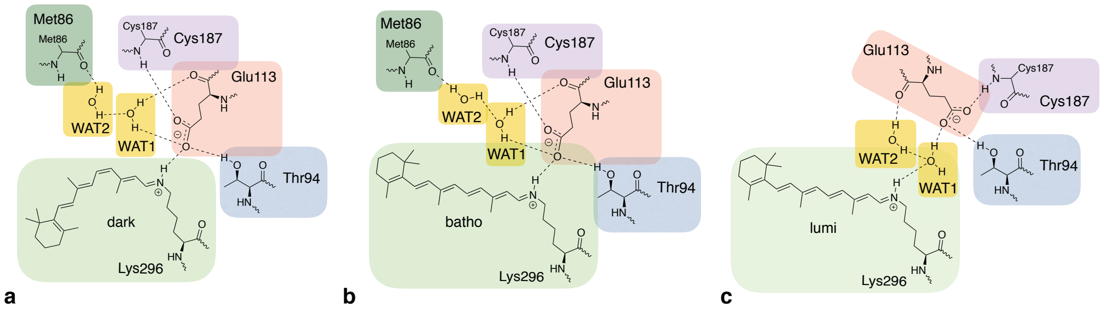


Figure 3.2: Difference in hydrogen-bond network around the protonated Schiff's base between the dark (a), batho (b) and lumi state (c). The hydrogen-bond network in the BSI state is the same as in bathorhodopsin.

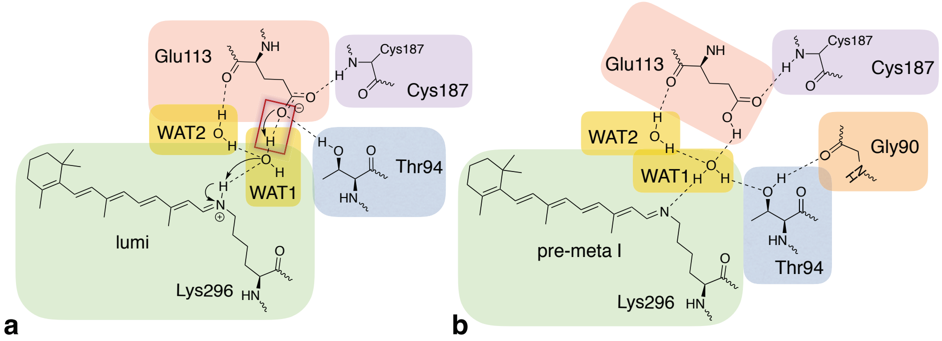


Figure 3.3: Conformational and configurational changes in the lumi state during and after deprotonation of retinal's positively charged nitrogen. (a) Deprotonation mechanism that takes placed during the TI calculations (Fig. 3.4). The red rectangle depicts the distance constraint between H(WAT1) and O1(Glu113) that was used during the TI simulations of the lumi state. (b) Change in the hydrogen-bond network after deprotonation of the protonated Schiff base in the lumi state in the last window (1.00 Å) of the TI QM/MM MD simulations and the unconstrained QM/MM MD simulation.

unconstrained simulation after the TI simulations have taken place (Fig. A.8, A.12, A.13). In the lumi state, however, the initial number of HBs of Glu113's O1 with the protein environment is two (Fig. 3.2), due to the presence of a bridging water (WAT1), which is decreased by one in the deprotonated lumi state or pre-meta-I configuration (Fig. 3.2 and Fig. 3.3). The pre-meta I state is a configuration obtained after 6 ps of unconstrained QM/MM MD post deprotonation

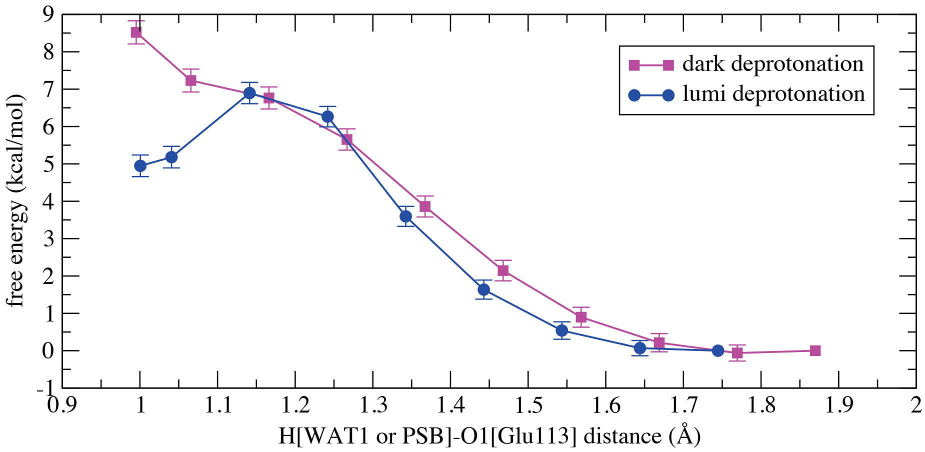


Figure 3.4: Free energy profile of the deprotonation of the chromophore in rhodopsin in the dark and lumi state. For the thermodynamic integration of the lumi state, the distance between one hydrogen of WAT1 and an oxygen of Glu113, O1, was used to perform the deprotonation reaction (see red rectangle in image 3.3*b*). In the dark state, the distance between the hydrogen on the nitrogen of the PSB, H(PSB), and Glu113's O1 oxygen was used as reaction coordinate in the TI simulations.

of the chromophore. The decrease in the number of HBs between the lumi and the pre-meta-I state is due to the reorientation of Thr94, which is initially interacting with Glu113, but forms a HB with the backbone oxygen of Gly90, Obb(Gly90), in the deprotonated form (Fig. 3.3, Fig. A.7, A.10, A.11).

Although the deprotonated retinal form is stable in the lumi state (Fig. A.7), the protonated Schiff base still has a lower free energy compared to the deprotonated form (Fig. 3.4). Therefore, thermodynamically, the deprotonated structure would still appear to be less stable than the protonated chromophore configuration. However, long-time scale rearrangements that occur upon proton transfer, which are not captured in the picosecond time scale QM/MM MD simulations, could further stabilise the deprotonated form. An approximate analysis tool, H++, was used to study the effect of the change in electrostatics of the active site upon deprotonation of the chromophore by calculating the pKa value of Glu113. Glu113, the glutamate that is protonated during deprotonation of the chromophore, instead of the chromophore itself was used as an indicator of the change in electrostatics, since H++ is unable to incorporate the charges of the chromophore into the pKa calculations. Nonetheless, the charge state of Glu113 can highly affect the pKa of the nearby retinal moiety [Sakmar et al., 1991] and could therefore be a measure of the changes in the electrostatic environment that favour/accompany the deprotonation reaction.

The pKa value of Glu113 has been calculated with several internal dielectric constants (Table

Table 3.1: pKa calculations of Glu113 in different rhodopsin intermediates. In addition, the pKa values for Asp83 and Glu122 are shown as reference values since these residues are believed to be protonated all throughout the activation [Fahmy et al., 1993]. pKa calculations were performed with internal dielectrics of 8, an external dielectric of 80 and a salinity of 0.15 M. The calculated shift is the total difference between the full (all) and the minimal (no Gly90, Thr94, WAT1, WAT2) model system, called *no G90, T94, waters*. Each time a molecule or residue is removed, the charges of the group are set to zero, but the radii are maintained. The full model includes rhodopsin without the chromophore and all water molecules present in the active site.

	dark			**lumi**			**pre-meta I**		
conformation	E113	D83	E122	E113	D83	E122	E113	D83	E122
all	4.7	10.9	8.9	4.8	10.7	8.7	8.5	11.2	7.8
no G90	4.5	11.0	8.9	4.8	10.8	8.7	8.1	11.3	8.1
no T94	5.7	10.9	8.9	5.8	10.7	8.7	8.4	11.2	8.0
no WAT1	5.5	10.8	8.9	5.9	10.6	8.7	7.5	11.2	8.2
no WAT1, WAT2	5.6	10.8	8.8	6.4	10.6	8.7	8.0	11.1	8.1
no T94, waters	6.5	10.8	8.9	7.4	10.6	8.6	7.7	11.1	8.2
no G90, T94, waters	6.3	11.0	8.9	7.3	10.8	8.7	7.3	11.2	8.3
shift	1.6	0.1	0.0	2.5	0.1	0.0	-1.2	0.0	0.5

3.1 and Table A.2) in several protein conformations and distributions of charge in the active site to provide a qualitative evaluation of the changes in the electrostatic environment when changing from the dark to the pre-meta-I conformation. In order to investigate the direct effect of the environment on Glu113, the presence/absence of the charge of all moieties that are hydrogen bonded to Glu113's O1 –the oxygen that is protonated during the chromophore's deprotonation– are taken into account (Fig. 3.2 and Fig. 3.3). Also the effect of Gly90 was studied because the residue substitutes Thr94's hydrogen-bond donor Glu113 and stabilises Thr94's change in orientation after deprotonation of the chromophore (Fig. 3.3).

The results in table 3.1 appear to suggest that although the pKa of Glu113 does not change significantly when moving from the dark state to the lumi conformation, the environmental dependence increases. In the dark state Glu113's pKa value is lowered by two units (mainly due to Thr94 and WAT1), while lumirhodopsin is lowered by three units (ca. one unit for Thr94, WAT1 and WAT2). Gly90 is not directly hydrogen bonded to Glu113 or a group in Glu113's HB network in the dark and lumi conformation and does not seem to have a significant effect on Glu113's pKa value. However, in pre-meta I, Gly90 appears to have an impact on the pKa value of Glu113, which could be due to the fact that upon deprotonation of the nitrogen of the PSB Gly90 is able to stabilise Thr94's change in orientation, leading to the removal of a hydrogen-bond donor from Glu113's HB network (Fig. 3.3).

While the change in the conformation of the active site from dark to lumirhodopsin is not reflected in the pKa value of Glu113 in the all conformation (Table 3.1), the nearby environment appears to have a strong effect on tuning the Glu113's pKa value when moving from the dark to the pre-meta-I state as the pKa value is increased by four units in pre-meta-I compared to the dark form. When removing charge from nearby moieties of Glu113, pre-meta I appears to be less affected than the dark state because pre-meta I is only increased by ca. one unit, mainly due to WAT1 and Gly90, slightly compensated by WAT2.

Hence, these pKa calculations appear to suggest that the protein environment has a significant effect on Glu113's protonation state. Moreover, after deprotonation, the HB network around Glu113 seems to cause Glu113's pKa to increase drastically, leading to values that favour a protonated form of Glu113 (Table 3.1, Table A.2).

3.4.2 Determination of Intermediates

The lumi and dark state structures, employed as initial configurations for the TI simulations described in Section 3.4.1, originated from snapshots taken from unrestrained classical MD simulations before and after *cis-trans* isomerisation, which were subsequently used to perform unconstrained QM/MM MD simulations. The QM/MM simulations of the lumi state were used to identify the lumi conformation via its optical and structural properties and lead to the selection of an initial structure for the investigation of the deprotonation reaction from which the pre-meta-I configuration resulted. The unconstrained QM/MM MD simulations of pre-meta I were generated after applying TI to the selected lumi state (see Section 3.4.1). Besides the dark, lumi and pre-meta-I conformations, additional snapshots of early intermediates were extracted from the unrestrained classical MD simulation after *cis-trans* isomerisation, to which also unconstrained QM/MM MD was applied in order to obtain simulations of the batho and the BSI state. All these QM/MM simulations of dark, batho, BSI, lumi and pre-meta I were used to generate configurations for the spectra calculations.

The calculated excitation energies of dark, batho, BSI and lumi closely reproduce the experimental data, which suggests that the obtained structures are good representatives of experimentally observed intermediates (Table 3.2). This conclusion is also supported by the alignment of the X-ray structures of the batho and lumi states with the QM/MM results (Fig. A.3, A.4). The obtained bathorhodopsin conformation is also in good agreement with previous computational studies [Campomanes et al., 2014; Lemaître et al., 2005]. Remarkable about the alignment of the lumi state (Fig. A.4) is that, while the overlap between the lumi structures is not as good as for the batho conformation (Fig. A.4), the space for a bridging water molecule is present in the X-ray structure, although the water molecule itself is absent in the experimental data. The absence of the water molecule could be due to the crystallisation techniques that were used to obtain the lumi form. The BSI configuration is more difficult to extract from the MD simulations as the configuration is in equilibrium with the batho state (Fig. A.5).

Table 3.2: Absorption spectra obtained via ZINDO/S for rhodopsin, bathorhodopsin, BSI, lumirhodopsin, pre-metarhodopsin I. ∗ 100 Configurations were used with a time interval of 12 fs to calculate the absorption maxima for dark, batho, lumi and pre-meta I rhodopsin. 44 Frames with a time interval of 12 fs were isolated for BSI. In the vertical-excitation energy calculations residues and water molecules closer than 7 Å to the retinal moiety were taken into account. Pre-metarhodopsin I represents the protein structure after deprotonation of the lumi state via QM/MM MD with TI. † The experimental absorption spectra were obtained at room temperature by Lewis *et al.* [Lewis et al., 2004].

	Absorption maxima (eV)		
Intermediates	ZINDO/S *	Exp. †	Difference
Rhodopsin	2.39 ± 0.15	2.48	0.09
Bathorhodopsin	2.17 ± 0.16	2.34	0.17
Blue-shifted intermediate	2.46 ± 0.12	2.60	0.14
Lumirhodopsin	2.34 ± 0.15	2.54	0.20
Pre-metarhodopsin I	2.94 ± 0.19	n/a	n/a
Metarhodopsin I	n/a	3.29	n/a

Besides the shifts in excitation energies also a modification of retinal's conformation can be clearly observed for the different states of rhodopsin. After *cis*-*trans* isomerisation, the methyl group on C13, C20, (Fig. 3.1a and Fig. 3.5) reorientates from a conformation that is more or less parallel to the membrane in the dark state to one that points towards the extracellular region in the lumi state (Fig. 3.5). The relaxation of the *trans* configuration via C20 is not surprising as the volume for the retinal moiety of the mainly hydrophobic/non-polar active site is tightly fitted around the β-ionone ring and the methyl group on C9 (C19) (Fig. 3.6). The only region in the active site that shows a relatively looser fit is the region around PSB, close to C20 (Fig. 3.6).

Together with the change in C20's location, the protonated nitrogen of Lys296 also alters its orientation in the active site due to retinal's change in conformation (Fig. 3.5). From the dark to the lumi state, the nitrogen moves from a position in which the atom points in a direction parallel to the membrane to a position in which it points more to the intracellular region. In addition to alterations in PSB's orientation, a shift in the hydrogen-bond network around PSB occurs, which takes place between the BSI and the lumi state. Consequently, batho and BSI contain a similar HB arrangement as the dark state (Fig. 3.2). The change in HB network in the lumi state implies the breakage of the direct salt bridge between PSB and Glu113 and the formation of an indirect hydrogen bond between Glu113 and PSB which is bridged by a water molecule (WAT1) (Fig. 3.2). The alteration in the hydrogen-bond network also contains a relocation of WAT2 and a decrease of hydrogen bonds from three to two for Glu113's O1 (Fig. 3.2). The changes in the hydrogen-bond network of pre-meta I in comparison to lumirhodopsin are shown in figure 3.3 and have been addressed in detail in Section 3.4.1.

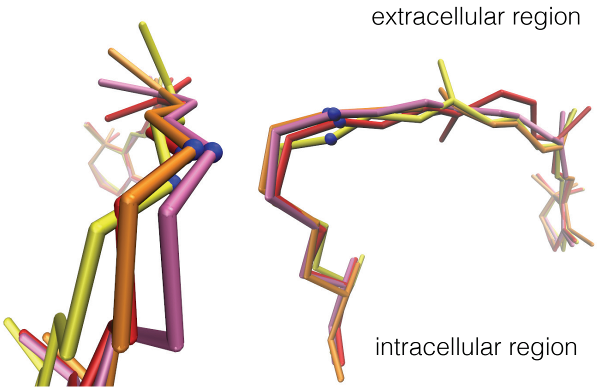


Figure 3.5: Overlay of rhodopsin structures in the dark (red) and three short-lived intermediate states: bathorhodopsin (pink), BSI (orange), lumirhodopsin (yellow). These structures were obtained via QM/MM MD and aligned on the retinal moiety and Lys296. The protonated nitrogen is highlighted through a blue ball representation.

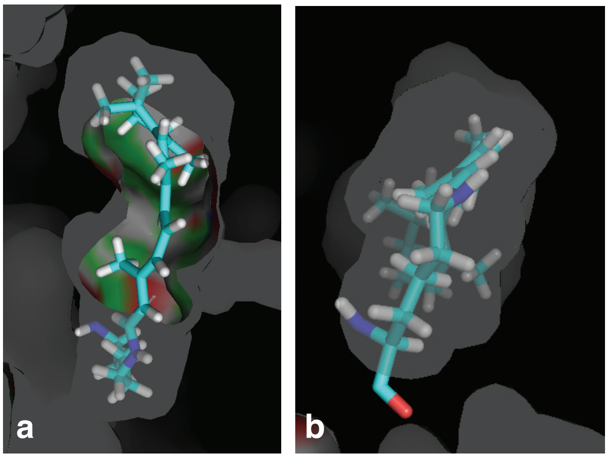


Figure 3.6: Volume of rhodopsin's active site showing the 11-*cis* configuration of the retinal moiety. (a) View of retinal moiety from the extracellular region. (b) View of the retinal moiety, turned 90° compared to image (a), with the Lys296 part located in front.

Although some structural rearrangements have already taken place in the pre-meta-I conformation and the optical properties have undergone a significant shift towards the meta I form, the chromophore and the protein surrounding are not yet fully relaxed (Table 3.2). While retinal's carbon chain has straightened compared to the lumi structure on the picosecond timescale, the β-ionone ring is still in the same location in the deprotonated *trans* conformation as prior to deprotonation (Fig. 3.7). This location of the β-ionone ring is not in line with experimental studies that show that the ring alters its location slightly in the meta I state due to a fully straightened conjugated chain compared to retinal's conformation in the early intermediates [Struts et al., 2011].

The conformational changes of the active site across the different intermediates is clearly evi-

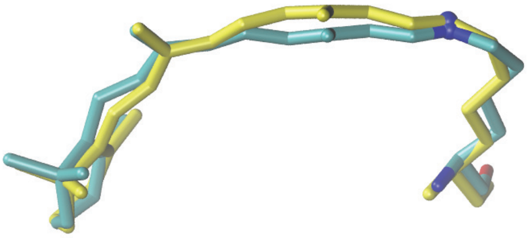


Figure 3.7: Overlay of the chromophore structures in the lumi state (yellow) and pre-meta-I configuration (cyan) viewed from the extracellular region. These snapshots were extracted from unconstrained runs of the configurations obtained before TI and as a result of the TI.

dent (Fig. 3.2). However, the interplay between the structural changes and the accompanying changes in the optical properties is not obvious. To identify the factors that are responsible for the spectral shifts of batho, lumi and pre-meta I, a correlation and causality inference analysis has been performed, which has already been applied by Campomanes *et al.* in a similar study [Campomanes et al., 2014]. This former study investigated the common factors that determine all shifts for the early intermediates up to lumirhodopsin by performing a single causality inference analysis over all snapshots from the included intermediates. Here, factors that influence individual shifts of the intermediates before deprotonation are studied as well as the pre-meta-I form. The BSI state was not taken into account because this state is in equilibrium with the batho state (Fig. A.5) and therefore only relatively few configurations can be unambiguously assigned to this intermediate.

The input for the correlation analysis of batho, lumi and pre-meta I included $48 * 10^6$ geometrical features per intermediate, which were composed of distances, angles, dihedrals and the BLA. The correlation profiles show that the $48 * 10^6$ features can be decreased to roughly 20 or 30 features having the highest heuristic merit by using Correlation-Based Feature Selection [Hall, 1999] (Fig. A.6). Remarkable is the fact that the shift for batho and lumi are defined by around 30 features, while the pre-meta I conformation (Fig. 3.3) only includes 19. Within the top ten of each intermediate, distance d(C9-C10) is present and the BLA is number one (Fig. A.6). The correlation between these two descriptors and the spectral shift has also been found in a computational study of rhodopsin by Campomanes *et al.* [Campomanes et al., 2014]. Distance d(C11-C12), the bond around which the *cis-trans* isomerisation takes place, appears to be correlated with the spectral shift in batho and lumi, but is replaced in the pre-meta I state by d(C12-C13) (Fig. A.6). This substitution of a double bond, d(C11-C12), for a single bond, d(C12-C13), in the causality of the spectral shifts could be caused by the deprotonation of the PSB that affects the properties of the conjugated chain. However, the origin of the bond exchange requires further investigation. Dihedrals are the most represented in the correlation profiles of the three intermediates, but they are not the same for different intermediates. Furthermore, a residue that has been reported to play an important role in rhodopsin's activation mechanism, Trp265 [Crocker et al., 2006; Standfuss et al., 2011], is also included in dihedrals and angles that describe the spectral shifts of batho, lumi and pre-meta I, showing that Trp265 is sensitive to the location of the retinal in the active site (Fig. A.6).

Investigating the inference of the causal relationship between features and the spectral shift provides a way to understand which features directly affect the spectral shift and which features only influence other features with a direct impact. The results for the three intermediates show that all intermediates are highly influenced by the BLA, pre-meta I>batho>lumi, as well as d(C9-C10), lumi>pre-meta I>batho (Fig. 3.8 and Table A.1). Similar to the correlation profiles, d(C11-C12) is substituted in the causality profile by d(C12-C13) when moving from lumi to pre-meta I. The causality between the spectral shift and the dihedrals is remarkable as well because, although specific dihedrals have a similar impact on the spectral shifts for each intermediate as the distances, the dihedral selection per intermediate is unique (Fig. 3.8). This could be due to the fact that the dihedrals are more sensitive to the orientation of the retinal and the conformation of the active site than the BLA and the distances. However, significant environmental alterations that take place from dark to pre-meta-I rhodopsin, such as the change in counter-ion distance, the rearrangement of water molecules in the active site or the protonation state of the chromophore, do not appear in the causality analysis or the correlation profiles (Fig. 3.8 and Fig. A.6).

Overall, the causality and correlation results appear to suggest that the descriptors with the most impact on the spectral shifts are intramolecular parameters such as the BLA and distances in the conjugated retinal chain. In addition, unique sets of dihedrals influence the optical properties of each intermediate.

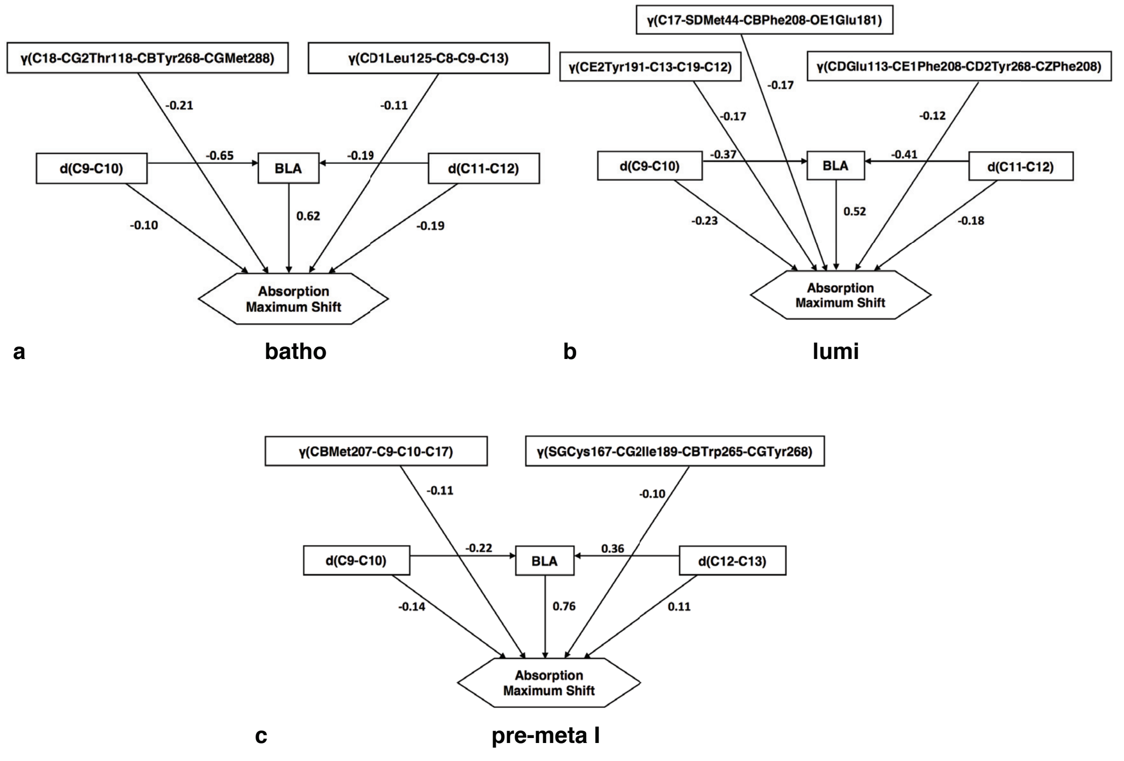


Figure 3.8: Markov blanket of the spectral shifts for the batho (a), lumi (b), and pre-meta I (c) configurations. The edge strength indicates the relative magnitude of the dependency between two variables, given the other inter-relationships.

3.5 Conclusions

The protonation/deprotonation equilibrium of rhodopsin's protonated Schiff base at body temperature is influenced by the conformational change of retinal after *cis*-*trans* isomerisation as well as the induced active site alterations, in particular the presence/absence of water molecules and accompanying rearrangements of the HB network. The first step in the relaxation pathway of the *trans* retinal group is governed by the anchoring of the β-ionone ring in the hydrophobic/non-polar pocket, the only available additional volume around C20 and the salt bridge between Glu113 and the chromophore. This step includes a clock-wise rotation of the C13-C14 bond, which induces the formation of the batho and BSI state, resulting in a C20 oriented towards the extracellular region together with a hydrogen-bond rearrangement near the PSB, leading to the lumi state. This HB rearrangement includes a water molecule that starts disrupting the salt bridge between Glu113 and the PSB in the lumi state as well as a decrease in hydrogen-bond donors for Glu113. Deprotonation in the lumi state results in a stable deprotonated form of the Schiff base, which neutralises the active site. As an effect of the proton transfer, further changes in the hydrogen-bond network take place that lead to a state in which the proton is stabilised on Glu113. Groups that appear to play a particularly important role in the deprotonation of PSB are two water molecules, WAT1 and WAT2, Gly90 and Thr94. The influence of the conformational rearrangements in the active site that occur during the transition from the dark to the pre-meta-I state is reflected by the performed pKa calculations of Glu113. While in the pKa calculations the effect of the environment is clear, the calculations of the spectral shift mainly show an impact of intramolecular descriptors, such as the BLA and the bond distances in the conjungated chain of the retinal moeity.

Further relaxation of the retinal moiety after deprotonation is required to reach the fully relaxed *trans* configuration of the meta I state as the shift between the absorption maxima of the experimentally observed meta I and the simulated early deprotonated form, pre-meta I, is significant. In a future study of the later intermediates in the photoactivation pathway, the relaxation of retinal towards the meta I state will be investigated by obtaining a new force field for the neutral *trans*-retinal configuration via the Force Matching method [Doemer et al., 2013] implemented in the CPMD package. This new force field will be employed in a classical MD simulation of rhodopsin in which *cis*-*trans* isomerisation and deprotonation have taken place to study the further relaxation of the chromophore.

4 Effect of N-Terminal Myristoylation on the Active Conformation of Gα_{i1}:GTP

Siri Camee van Keulen, Ursula Röthlisberger

4.1 Summary

G proteins are part of the G-protein coupled receptor (GPCR) signal-transduction cascade in which they transfer a signal from the membrane-embedded GPCR to other proteins in the cell. In case of the inhibitory G-protein heterotrimer, permanent N-terminal myristoylation can transiently localise the Gα_i subunit at the membrane as well as crucially influence Gα_i's function in the GTP-bound conformation. The attachment of lipids to proteins is known to be essential for membrane trafficking, however, our results suggest that lipidation is also important for protein-protein interactions during signal transduction. Here we investigate the effect of myristoylation on the structure and dynamics of soluble Gα_{i1} and its possible implication for signal transduction. A 2 μs classical molecular dynamics simulation of a myristoylated Gα_{i1}:GTP complex suggests that the myristoyl-induced conformational changes of the switch II and alpha helical domains create new possibilities for protein-protein interactions and emphasise the importance of permanent lipid attachment for the conformation and functional tunability of signalling proteins.

4.2 Introduction

G-protein-coupled receptor (GPCR) pathways are of high interest since their signal-transduction cascades play an important role in several diseases such as heart failure, hypertension and obesity. The first proteins that will transfer an extracellular signal, received by a GPCR, to other regions in the cell, are G-proteins, which are therefore crucial messengers in GPCR regulated signalling. G proteins are built from an α, β and γ subunit of which the α subunit is able to interact with guanosine diphosphate (GDP) in the protein's inactive state or with guanosine-5'-triphosphate (GTP) in the active conformation. The interaction between a heterotrimeric G protein and an active GPCR induces GDP/GTP exchange in the Gα subunit [Dror et al., 2015; Louet et al., 2012; Schröter et al., 2015]. A concurrent destabilisation of the interface between Gα and the G$\beta\gamma$ dimer [Galés et al., 2006] leads to soluble Gα, thereby decreasing its concentration close to the membrane, in favour of the one in the cytosol [Preininger et al., 2012]. The specificity by which G-protein subunits interact with receptors and effectors defines the variety of responses that a cell is capable of providing due to an extracellular signal. Hence, understanding the molecular mechanism of G proteins' binding to their cellular partners is a key step in understanding GPCR signalling cascades and can shed more light on the malfunctioning of GPCR signalling in humans.

Similarly to G proteins, smaller monomeric GTPases such as the ADP-ribosylation factor (ARF) proteins are also activated by GDP/GTP exchange. The localisation of ARF proteins, like ARF1, at the membrane depends on whether the protein is interacting with a GDP or a GTP molecule [Liu et al., 2009, 2010]. In addition, ARF1's location is not only determined by nucleotide interaction but also by lipidation of the N-terminus through myristoylation. The myristoyl group interacts with the membrane in ARF1's active conformation while it is interacting with a hydrophobic pocket on the protein in the inactive soluble state. Hence, GDP/GTP exchange and the myristoyl moiety are important for membrane localisation and the solubility of ARF1 [Behnia and Munro, 2005]. In the case of G proteins, similar factors can be found that influence the localisation of Gα subunits. Besides GDP/GTP exchange, also lipidation plays a role in Gα localisation. All Gα families (Gα_s (stimulatory), Gα_i (inhibitory), Gα_q, Gα_{12}) can reversibly bind palmitoyl groups to the N-terminus of their α subunit (except for Gα_t from the Gα_i family), which anchors the subunit to the membrane, but only the Gα_i family has a myristoyl group attached to its N-terminus in addition to the palmitoyl moiety [Chen and Manning, 2001; Cabrera-Vera et al., 2003]. Unlike palmitoylation, myristoylation only enables transient binding to the membrane, which can lead to an equilibrium between a membrane-bound and a solvated state [Wright et al., 2010]. Moreover, myristoylation is an irreversible process which results in myristoyl attachment throughout the lifetime of the protein [Cabrera-Vera et al., 2003]. Hence, the increased solubility of all Gα families' active Gα:GTP complex is achieved by a combination of (i) dissociation of the G-protein heterotrimer, (ii) depalmitoylation [Loisel et al., 1999; Wedegaertner and Bourne, 1994] and (iii) detachment of the α-subunit N-terminal helix from the lipid bilayer. The solubility of the Gα_i family is also influenced by myristoylation as the post translational modification induces an equilibrium between membrane-bound and cytosolic forms [Degtyarev et al., 1994].

Besides the solubility, also the conformation of active Gα_{i1} has been shown to be affected by myristoylation [Preininger et al., 2003], as the soluble non-myristoylated Gα_{i1}:GTP complex does not maintain the secondary structure of its N-terminal helix, while the active soluble myristoylated Gα_{i1} subunit shows a more structured N-terminal tail and is able to position the N-terminus close to the Gα_{i1} subunit [Preininger et al., 2003]. The importance of permanent N-myristoylation of the Gα_{i1} subunit is not only shown by its influence on Gα_{i1}:GTP's structure but also by the dependence of the function of the α subunit on the presence of the myristoyl moiety [Dessauer et al., 1998; Taussig et al., 1993]. In the case of adenylyl cyclase type 5 (AC5), for example, myristoylated Gα_{i1} (Gα_{i1}^{myr}) can interact with the catalytic domain of AC5 and inhibit its catalytic function. However, a non-myristoylated Gα_{i1} (Gα_{i1}^{non}) subunit is incapable of interacting with AC and can therefore not inhibit AC's ATP conversion pathway [Dessauer et al., 1998]. For the stimulatory Gα_s though, the presence of the N-terminus is not crucial for its interaction with AC as is demonstrated by the obtained X-ray structure of Tesmer *et al.* [Tesmer et al., 1999]. It can be proposed that in the active Gα_s state, the N-terminus is less structured, similar to the results of Preininger *et al.* for the non-myristoylated Gα_{i1}:GTP complex, which could suggest that the overall active Gα_s conformation is not significantly altered by the structure of the N-terminal helix [Preininger et al., 2003].

Although myristoylation appears to be an important difference between the stimulating Gα_s and inhibiting Gα_i families and is furthermore crucial for Gα_{i1}'s conformation, no X-ray structures of myristoylated Gα_i proteins are currently available in the Protein Data Bank (PDB) [Berman et al., 2000]. The available X-ray structures of active Gα_s and Gα_{i1}^{non} are similar as is shown via a root mean square deviation (RMSD) of 1.07 Å between the two structures (Fig. B.1). Especially around the proposed protein interaction site, the switch II region, the subunits align well, while their function is inverse in the cytoplasmatic domain. Therefore, it is unclear to what extent the available Gα_{i1}^{non} structures are representative of the complete conformational range of the active Gα_{i1} protein. Hence, understanding the effects of myristoylation on Gα_{i1} can possibly help rationalise the impact on Gα_{i1}'s conformation and the difference in biological function between Gα_s and Gα_{i1}. Additionally, it can contribute to a better understanding of G-protein activation and provide a new perspective on the role of lipid attachment in protein function.

To shed light on this issue, we have studied Gα_{i1}^{myr} protein in its active state using classical molecular dynamics (MD) simulations of a Gα_{i1}^{myr} model (Fig. 4.3) based on Gα_{i1}^{non} X-ray structures and an ARF1 protein NMR structure (Fig. 4.1 and see Section 4.3.1). A second simulation was performed as well with the same Gα_{i1}:GTP model in the absence of its myristoylated

N-terminus as a cross validation.

A 2 μs simulation of the Gα_{i1}^{myr}:GTP complex, identifies a stable binding site for the myristoyl group in between α1, α5 and β2-β3, and demonstrates the effect of myristoyl binding for important interaction sites such as the switch II region and the nucleotide binding pocket. Significant changes occur in the conformation of the switch II region upon incorporation of the myristoyl moiety. Moreover, a large rearrangement of αB in the alpha helical (AH) domain of the α subunit takes place. These changes in the secondary structure of the switch II region and the AH domain could have an important effect on the way the Gα_{i1}^{myr} subunit is able to interact with proteins in the cytosol. The proposed mechanism for G$_{i1}^{myr}$ activation is described in figure 4.2.

4.3 Methods

4.3.1 Homology Modelling

The refinement protocol automodel from Modeller 9.14 [Eswar et al., 2006; Martí-Renom et al., 2000] was employed to obtain the initial homology model of the myristoylated Rattus norvegicus Gα_{i1} subunit (UniprotKB P10824) that was used to start the classical MD runs (Fig. 4.3a and Fig. B.2). In total 100 structures were obtained with Modeller 9.14. Besides maintaining the secondary structure of the main template, 4PAQ, and avoiding bad contacts within the structure, the selection of the initial model for simulation was based on (I) the presence of important hydrogen bonds between GTP and the active site that are shown in the X-ray structure, (II) maintaining the secondary structure present in the included templates

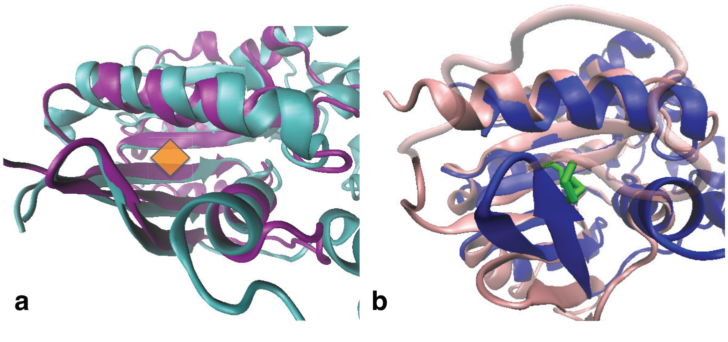
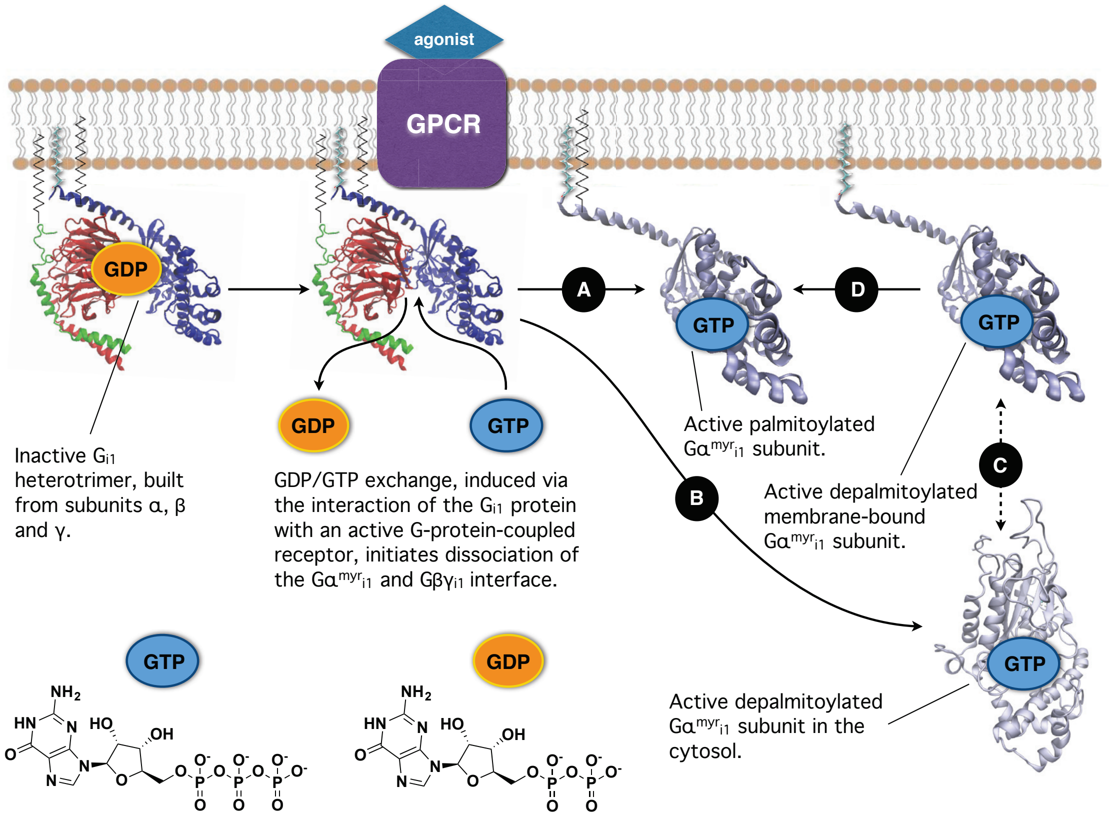


Figure 4.1: Comparison of ARF1 and Gα_{i1}. (a) Alignment of the Cα atoms of the ARF1:GTP complex in *purple* (PDB code 2KSQ) and the Gα_{i1}^{non}:GDP conformation in *cyan* (PDB code 1AS3) zoomed in on ARF1's myristoyl binding site (*orange* diamond) in the inactive conformation. (b) Alignment of the Cα atoms of the ARF1:GDP complex in *pink* (PDB code 2K5U) and the Gα_{i1}^{non}:GTP conformation in *blue* (PDB code 1GIA) zoomed in on ARF1's myristoyl binding site. The myristoyl moiety of ARF1 is depicted in *green* and is interacting with a hydrophobic pocket, located on ARF1.



Inactive G_{i1} heterotrimer, built from subunits α, β and γ.

GDP/GTP exchange, induced via the interaction of the G_{i1} protein with an active G-protein-coupled receptor, initiates dissociation of the $G\alpha^{myr}_{i1}$ and $G\beta\gamma_{i1}$ interface.

Active palmitoylated $G\alpha^{myr}_{i1}$ subunit.

Active depalmitoylated membrane-bound $G\alpha^{myr}_{i1}$ subunit.

Active depalmitoylated $G\alpha^{myr}_{i1}$ subunit in the cytosol.

Figure 4.2: Proposed activation mechanism of G^{myr}_{i1} in which an activated GPCR induces GDP/GTP exchange that leads to dissociation of the $G\alpha_{i1}$ and $G\beta\gamma_{i1}$ interface. *A* represents the activation of $G\alpha^{myr}_{i1}$ in which the N-terminal palmitoyl group is still present. *B* shows the conversion to activated depalmitoylated $G\alpha^{myr}_{i1}$ which results in an equilibrium (*C*) between a solvated $G\alpha^{myr}_{i1}$:GTP and a membrane-bound $G\alpha^{myr}_{i1}$:GTP complex. *D* represents the palmitoylation of active membrane-bound $G\alpha^{myr}_{i1}$.

of the regions that were not part of the main template, PDB code 4PAQ, (III) the position of the myristoyl moiety that had to be located within 4 Å of the hydrophobic pocket as an initial guess for the moiety's location.

A combination of three different X-ray structures (PDB codes: 2K5U, 1AS3 and 4PAQ) were incorporated in the initial homology model [Kaya et al., 2014; Liu et al., 2009; Raw et al., 1997]. 4PAQ, a Rattus norvegicus Gα_{i1}^{non}:5'-guanosine-diphosphate-monothiophosphate (GTPγS) complex, was used as a main template. GTPγS is an analog of GTP, which is used in the crystallisation process of the Gα_{i1} subunit as GTPγS is not able to undergo the hydrolysis mechanism in the active site of the protein. 1AS3, a Rattus norvegicus Gα_{i1}^{non}:GDP conformation, covered the N (residues 9-34) and the C (residues 328-354) terminus of the Gα_{i1} subunit. In order to obtain an initial guess for the location of the myristoyl moiety, a myristoyl conformation of ARF1 from saccharomyces cerevisiae (UniprotKB P11076) with PDB code 2K5U was used (myristoyl group and residues 1-7). The 2K5U NMR structure was also employed to model the interswitch toggle (see Section 4.4 and *Supporting Methods* in Appendix B), residues 179-220 of Gα_{i1}, which enables communication between the myristoyl and the nucleotide binding site [Pasqualato et al., 2002].

The location of GTPγS and the Mg^{2+} ion were taken from 4PAQ in order to place the ion and guide the GTP molecule in the nucleotide binding site. Also water molecules (16 molecules in total) surrounding GTPγS in 4PAQ were incorporated in the model. Since different structures were used to obtain the homology model, restraints were employed for the construction of the initial model to maintain important hydrogen bonds in the nucleotide binding site and to position the myristoyl moiety close to the hydrophobic pocket (Fig. B.2).

4.3.2 Classical Molecular Dynamics Simulations

The Gα_{i1} together with its bound myristoyl group, Mg^{2+} ion and GTP were simulated for 2 μs at 310 K and 1 bar using a Nosé-Hoover thermostat and an isotropic Parrinello-Rahman barostat. Additionally, about 42 000 water molecules and 150 mM KCl are present in the simulated system. The force fields used for the protein and the water molecules are AMBER99SB [Hornak et al., 2006] and TIP3P [Jorgensen et al., 1983], which were employed by Gromacs 4.6.6 [Bekker et al., 1993; Hess et al., 2008] to perform the runs. For GTP, the force field parameters generated by Meagher *et al.* were used [Meagher et al., 2003]. The adjusted force field parameters for K^+ and Cl^- were taken from Joung *et al.* [Joung and Cheatham III, 2008]. The Mg^{2+} parameters originated from Allnér *et al.* [Allnér et al., 2012]. The charges for the myristoyl group were obtained with Gaussian 09 [Frisch et al.] based on Hartree Fock calculations in combination with a 6-31G* basis set and using the AMBER RESP procedure [Case et al., 2015] (Fig. B.4). Appropriate atom types from the AMBER99SB force field were selected to complete the myristoyl description.

Electrostatic interactions were calculated with the Ewald particle mesh method with a real space cutoff of 12 Å. Bonds involving hydrogen atoms were constrained using the LINCS algorithm. The time integration step was set to 2 fs.

4.3.3 Calculation of Electrostatic Surfaces

The electrostatic surface of several Gα_{i1} subunits was calculated via APBS 1.3 [Baker et al., 2001; Unni et al., 2011]. The multigrid method together with the monopole boundary conditions and spline-based discretisation of the delta functions were used. The ion concentration was set to 150 mM at a temperature of 310 K. The default settings for the biomolecular (2.00) and solvent (78.54) dielectric constants were used. PDB2PQR 2.0.0 [Dolinsky et al., 2007; Unni et al., 2011] was employed to perform the preprocessing of the conformations with the AMBER force field in order to calculate the atomic radii, charges and position of the hydrogen atoms.

4.3.4 Structure Superpositions and Images

Multiprot [Shatsky et al., 2004] and VMD [Humphrey et al., 1996] were used to align protein structures. The Align tool in Uniprot [Consortium, 2014] was used to align protein sequences. Images were prepared with VMD [Humphrey et al., 1996].

4.4 Results and Discussion

4.4.1 Classical Molecular Dynamics Simulations of Gα_{i1}^{myr}

During the 2 μs of classical MD simulation, the effects of the myristoyl group on the structure of the active Gα_{i1} subunit become apparent (Fig. 4.3 and Fig. B.5). In the equilibrated conformation, the covalently-bound myristoyl group interacts with the Gα_{i1} protein via β2-β3, the α5 helix and the α1 helix (Fig. 4.3b-4.3d). One direct effect of the presence of the myristoyl group is an outward conformation of β2-β3 in order to provide space for the lipid moiety (Fig. 4.4a). The β2-β3 strands are shifted down by two residue units (Fig. 4.4b), which increases the number of residues that are part of the switch II region. This shift of two residue units has also been observed in myristoyl containing ARF proteins and is called the interswitch toggle [Pasqualato et al., 2002]. This is not too surprising since ARF's signature sequence for the interswitch toggle, wDvGGqxxxRxxW, is also present in Gα_{i1}'s sequence in the region that includes β3 and the switch II region (Fig. 4.5) [Pasqualato et al., 2002]. The interswitch toggle facilitates communication within the Gα_{i1} subunit between the nucleotide binding site and the myristoyl moiety [Pasqualato et al., 2002]. The effect on the nucleotide binding site (see *Supporting Results and Discussion* in Appendix B) is that since there are more residues available from the switch II region, the N-terminal region of the switch II region is able to

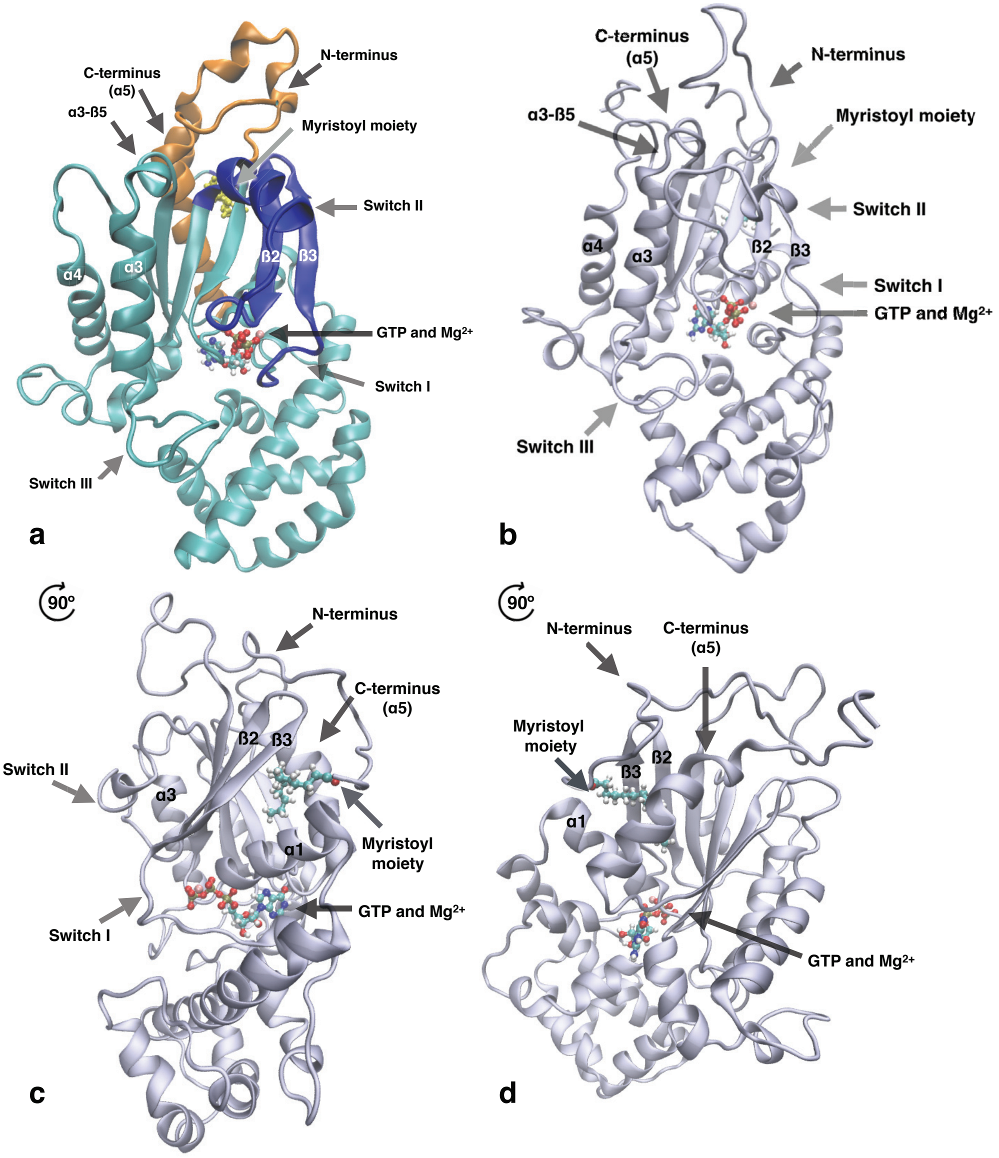


Figure 4.3: Structural overview of $G\alpha_{i1}^{myr}$:GTP and comparison with X-ray structures. (a) Representation of the initial homology model that was used to start the classical MD simulation. The template for the *cyan* region was PDB entry 4PAQ, for the *orange* area the PDB structure 1AS3 was used and the *blue* part was modelled after PDB entry 2K5U. The myristoyl moiety is depicted in *yellow.* (b, c, d) Three different views of the final $G\alpha_{i1}^{myr}$ structure in which the myristoyl moiety, GTP and Mg^{2+} are also shown.

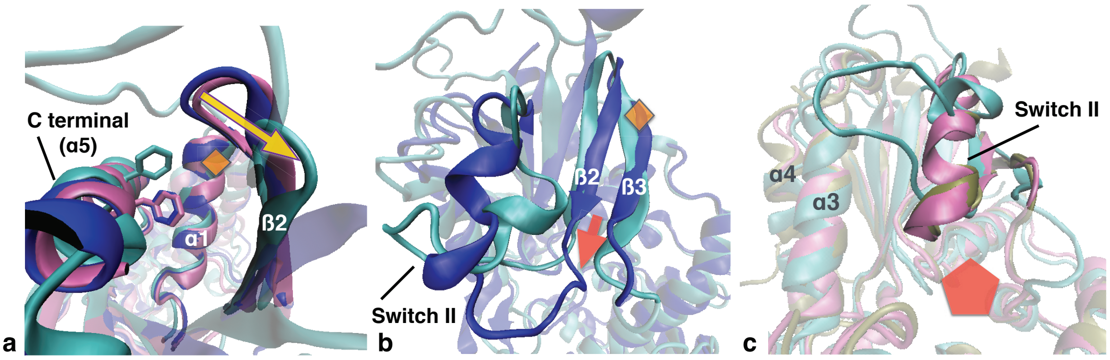


Figure 4.4: (a) Zoom on the proposed myristoyl binding site of aligned Gα_{i1} subunits interacting with GTP, a GTP analog or GDP. The *mauve* coloured structure is a Gα_{i1}^{non}:GTPγS complex (PDB code 1GIA) and the model, Gα_{i1}^{myr}:GTP, after about ~ 1.8 μs is shown in *cyan*. The G$_{i1}^{non}$:GDP conformation (PDB code 1GP2) is coloured *blue*. By the *orange* diamond it is shown where the myristoyl group is connected to the Gα_{i1} subunit of the model. The *yellow* arrow specifies the outward movement of the β2-β3 domain upon myristoyl binding. Additionally, Phe336's orientation is shown for all three structures in the previously described structure colours. (b) The β2-β3 shift between the G$_{i1}^{non}$:GDP and the Gα_{i1}^{myr}:GTP complexes is shown by the *red* arrow and the myristoyl binding site is depicted as an *orange* diamond. The color scheme for the structures is the same as in the *a* image. (c) Alignment of Gα_{i1} subunits in which the difference in the switch II region is shown. The *red* pentagon is showing the location of the guanine nucleotide binding site. The colour scheme of the Gα_{i1} structures is the same as in the *a* image, except for the fact that the *tan* structure is the X-ray conformation of the Gα_s:GTPγS complex (PDB code 1AZT) [Sunahara et al., 1997b].

interact with the α3 helix, which stabilises the structure of the switch II region (Fig. 4.4c). A cross-validation simulation was also performed in which the N-terminus was cut off until the 31th residue, Gα_{i1}^{non}:GTP, to monitor the effect on the initial model (Fig. B.6). The result showed indeed a hydrophobic pocket that was closed by β2-β3 and a switch II region that returned to the conformation of the switch II region of the Gα_{i1}^{non}:GTPγS X-ray structure (PDB code 1GIA). The upward movement of the interswitch toggle and the full rearrangement of the nucleotide binding site was not observed in the 1 μs classical MD simulation, which could be due to the limited length of the simulation. Overall, the structural rearrangements of the simulated Gα_{i1}^{myr} as a result of myristoyl binding, show that the effect of the myristoyl moiety on the conformation of the Gα_{i1}^{myr} subunit is not limited to the hydrophobic binding pocket

```
P11076 ARF1_YEAST    56  VQYKNISFTVWDVGGQDRIRSLWRHYYRNTEGVIFVVDSNDRSRI 100
P08754 GNAI3_HUMAN  189  FTFKDLYFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLV 233
B2RSH2 GNAI1_MOUSE  189  FTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLV 233
P10824 GNAI1_RAT    189  FTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLV 233
P50146 GNAI1_CHICK  189  FTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLV 233
P63096 GNAI1_HUMAN  189  FTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLV 233
P04899 GNAI2_HUMAN  190  FTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSAYDLV 234
                         .  ..: * ::*****   *  * : :. . .:** *  .  . :
```

Figure 4.5: Alignment of several Gα sequences and ARF1 from saccharomyces cerevisiae.

but influences also other parts of the protein.

The change in conformation of Gα_{i1}^{myr}:GTP is likely to also have implications for the interaction with other proteins, which could affect the overall function of Gα_{i1}^{myr} in the cytosol. In order to address the role of the altered conformation and to identify if it has an impact on protein-protein interactions in which active Gα_{i1} is involved, the equilibrated model was compared to X-ray structures of Gα_{i1}^{non} and Gα_s. First, the overall changes in conformation are compared and a mechanism for Gα_{i1}^{myr} localisation in the cytosol is proposed. Moreover, the important regions for myristoyl binding, GTP binding (see *Supporting Results and Discussion* in Appendix B), and protein-protein interactions are discussed to obtain a better understanding of the effect of permanent lipid attachment.

4.4.2 Two Conformations of Active Gα_{i1}^{myr}

When comparing the simulated model with the crystal structure of the Gα_{i1}^{non}:GTPγS complex (PDB code 1GIA) [Coleman et al., 1994] in which the N-terminus is absent, it becomes evident that in the myristoylated Gα_{i1}:GTP complex an outward orientation of β2-β3 is stabilised by the myristoyl in the hydrophobic pocket. Because of this change in conformation also the conformation of the switch II region is modified. In the X-ray structure of Gα_{i1}^{non}:GTPγS, however, the conformation of the complex is strikingly identical to the active Gα_s:GTPγS conformation as the myristoyl binding pocket remains closed (Fig. 4.4c and Fig. B.1) [Coleman et al., 1994; Sunahara et al., 1997b]. Although in both X-ray structures, Gα_s:GTPγS and Gα_{i1}^{non}:GTPγS, the N-terminus is not present, the N-terminus of Gα_s:GTP is suggested to be less structured in solution as it is not anchored on the protein by a myristoyl group or by the membrane, while in the case of Gα_{i1}^{myr}:GTP, the N-terminus adopts a more structured conformation in close proximity to the protein [Preininger et al., 2003]. This difference in myristoyl presence could suggest that the impact of the missing N-terminus is less severe for the conformation of the protein in the case of Gα_s:GTPγS than the one of Gα_{i1}:GTPγS [Preininger et al., 2003].

The X-ray structure of activated Gα_{i1}^{non} and the simulated active Gα_{i1}^{myr} appear to represent two different conformations of Gα_{i1} in the active state. The main difference between the two structures is the absence or presence of the myristoylated N-terminus. Possibly, both conformations are able to exist under two different circumstances: one in which the myristoyl group is interacting with the protein's hydrophobic binding pocket and one in which it is immersed into the membrane, localising Gα_{i1} in close proximity to the membrane surface. This suggestion is consistent with experimental findings in which both membrane-bound and solvated conformations are simultaneously observed for active Gα_{i1}^{myr} [Degtyarev et al., 1994]. Consequently, in this view two forms of activated depalmitoylated Gα_{i1}^{myr} are present in equilibrium in the intracellular domain: the membrane-bound and the solvated form of Gα_{i1}^{myr}:GTP (Fig. 4.2). This also proposes that palmitoylation in combination with myristoylation could be responsible for different conformations of the Gα_{i1}:GTP complex by transiently or permanently binding the subunit to the lipid bilayer. This directionality induced by lipid attachment could supposedly result in selectivity for particular proteins, depending on Gα_{i1}^{myr}'s location in the cell.

The X-ray structure of Gα_{i1}^{non}:GTPγS is likely to represent the membrane-bound-like conformation of active Gα_{i1}^{myr} as the hydrophobic pocket for the myristoyl moiety, that is identified by the classical MD simulation to be between β2-β3, the α5 helix and the α1 helix, is unavailable. Consequently, the lipid will have to be located in a different hydrophobic site, most probably at the membrane itself [Linder et al., 1991]. It is known from experiments that the myristoyl group alone is not enough to locate Gα_{i1} at the membrane permanently but enables a transient stability which results in a fraction of the protein being in the membrane-bound state and another fraction in the solvated state [Degtyarev et al., 1994]. In fact, in order to stabilise its location on the membrane in a definitive manner, Gα_{i1}^{myr} has to be palmitoylated, which results in a firm anchoring of the protein to the lipid bilayer [Chen and Manning, 2001; Degtyarev et al., 1994; Galbiati et al., 1996]. In view of this, the active conformation of Gα_{i1}^{myr} observed in the simulations appears to be a structural representative of solvated depalmitoylated Gα_{i1}^{myr}:GTP with the myristoyl group packed away in a hydrophobic protein pocket (see Section 4.4.3).

In the next sections, the Gα_{i1}^{myr}:GTP structure in the solvated state will be described in more detail and compared to the Gα_{i1}^{non}:GTPγS conformation.

4.4.3 Myristoyl Binding Site

Conformational changes induced by the myristoyl moiety start with the modification of Gα_{i1}^{myr}'s hydrophobic pocket (Fig. 4.6a, 4.6b). Within the time scale of the 2 μs simulation, the hydrophobic site in the Gα_{i1}^{myr}:GTP complex assumes a more stable conformation. β2-β3 gets closer to α1 and the myristoyl tail reorients (Fig. B.7) while the region of the hydrophobic pocket remains similar to the ARF1:GDP complex (Fig. 4.6c). The location of the myristoyl moiety in the pocket is close to the position of the myristoyl group in ARF1, but the orientation of the myristoyl tail differs between Gα_{i1}^{myr}:GTP and ARF1:GDP. In the ARF1:GDP conformation the majority of the myristoyl is positioned between β2-β3 and α5, while the tail is pointing towards the C-terminal end of α5. In the Gα_{i1}^{myr}:GTP system, the myristoyl group is in between β2-β3 and α5 as well, but the tail is oriented towards α1 (Fig. 4.6c). These alterations are not surprising in view of the fact that the changes in the hydrophobic pocket conformation could originate from the difference in the *Ras* domain between the two proteins. Gα_{i1}, for instance, has a longer α1 helix than ARF1, that can stabilise β2-β3. Additionally, an alpha helical domain is located between α1 and switch I in the Gα_{i1} protein, which restrains the conformational flexibility of switch I in comparison to ARF1.

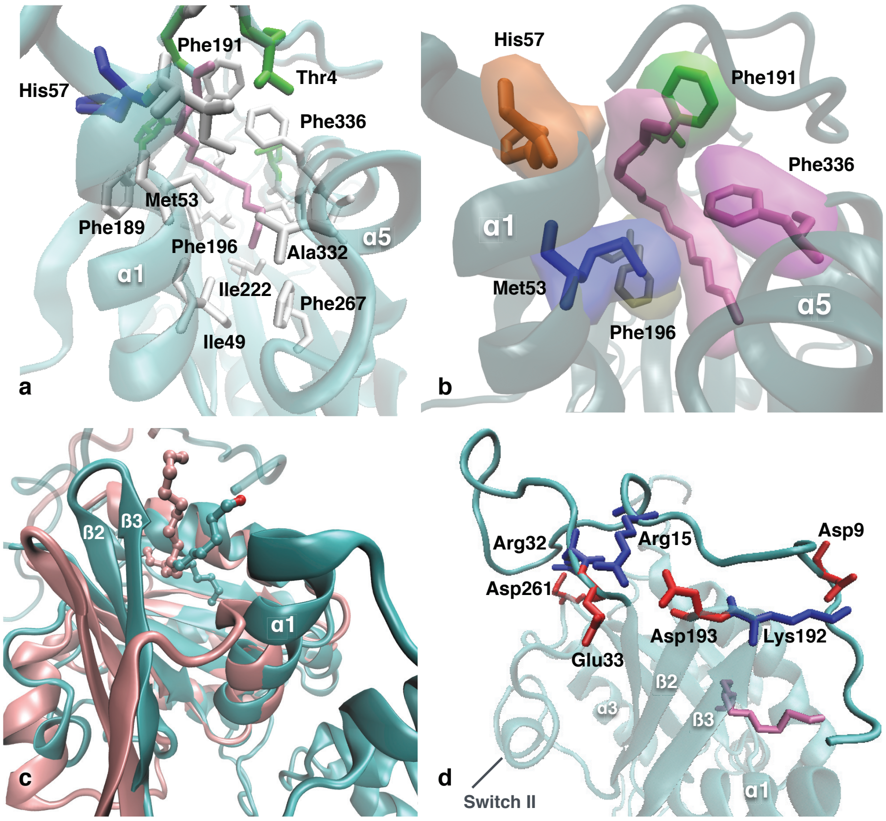


Figure 4.6: Detailed representation of Gα_{i1}^{myr}'s myristoyl binding site and N-terminus. (a) Zoom on Gα_{i1}^{myr}'s hydrophobic binding pocket in which the myristoyl group (*pink*) is binding. Non-polar residues are colored in *white*, polar residues are shown in *green*. (b) Myristoyl binding pocket of the simulated Gα_{i1}^{myr}:GTP complex in which the myristoyl moiety (*pink*) and several residues are shown: Met53 (*blue*), His57 (*orange*), Phe189 (*green*), Phe191 (*gray*) and Phe336 (*purple*). In addition, the molecular volume sampled over a period of 500 ns, from 1.6 to 2.1 μs, is shown for all described moieties. (c) Alignment of Gα_{i1}^{myr}:GTP in the equilibrated conformation (*cyan*) at ~ 1.8 μs of classical MD and ARF1:GDP (*pink*), PDB code 2K5U. Besides the protein, also the location of the myristoyl group, GTP , GDP and Mg^{2+} are shown for both conformations. (d) Location of the N-terminus with respect to the rest of the Gα_{i1}^{myr} subunit. The myristoyl moiety is shown in *pink*. Negatively (*red*) and positively (*blue*) charged residues are shown on and around the N-terminus that keep the N-terminal tail into place on the Gα subunit.

Several features of the hydrophobic pocket of Gα_{i1}^{myr}:GTP can be rationalised through comparison with experimental findings (see *Supporting Results and Discussion* in Appendix B). For instance, due to the location of the myristoyl group, an outward motion of Phe336 occurs during the classical MD simulation (Fig. 4.4a), which has been linked to activation of the Gα_{i1} subunit [Kaya et al., 2014] (see *Supporting Results and Discussion* in Appendix B).The identified hydrophobic pocket also opens up a new region for the investigation of myristoyl binding to Gα_{i1}. The presented data show that residues Met53 and Phe196, for instance, are in close proximity to the myristoyl group and therefore these residues could be candidates for mutagenesis studies as mutating these residues into a bulky hydrophobic residue, such as a tryptophan, could prohibit the myristoyl moiety from entering the hydrophobic pocket. Via these mutations a change in the conformation of the Gα_{i1} subunit could be initiated, resulting in a hydrophobic pocket that is unavailable for the myristoyl moiety. This mutated α subunit would be myristoylated, however, it would be unable to interact with adenylyl cyclase type 5 due to the change in the hydrophobic pocket.

Besides the hydrophobic pocket, the conformation of the N-terminus is also adjusted by the presence of the myristoyl group (see *Supporting Results and Discussion* in Appendix B), which becomes anchored through the embedding of the myristoyl moiety in the hydrophobic binding pocket. In the Gα_{i1}^{myr}:GTP complex, the N-terminus is closely packed to the rest of the protein and is not only able to interact with β2-β3 and α3-β5, stabilising the conformation of the hydrophobic pocket, but can also form interactions with the C-terminal part of the switch II region (Fig. 4.6d). The presence of the myristoyl moiety anchors the N-terminus in a more rigid conformation that enables further stabilisation of the hydrophobic pocket.

4.4.4 Switch II Region and Alpha Helical Domain

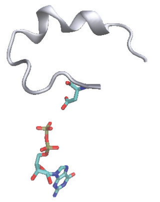
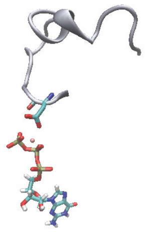
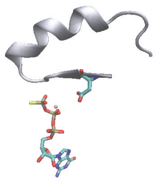
Available X-ray structures and the data from the classical MD propose at least two possible locations of the myristoyl moiety during different states of the Gα_{i1} subunit (Table 4.1). The Gα_{i1}^{non}:GDP complex, in which the majority of the N-terminus is resolved, suggests that the myristoyl group is located near the membrane and is not able to interact with the hydrophobic pocket. The Gα_{i1}^{non}:GTPγS structure is a conformation in which the entire N-terminus is absent and could therefore be perceived as a conformation of active Gα_{i1} with the myristoyl group not docked in the hydrophobic pocket but in a different hydrophobic region, such as the membrane. When Gα_{i1}^{non}:GTPγS, Gα_{i1}^{myr}:GTP and Gα_{i1}^{non}:GDP are compared, the location of the myristoyl appears to affect also other regions of the subunit than the hydrophobic binding pocket, such as the switch II region (Table 4.1).

The differences in the switch II region between the three states is evident. A unique feature of the solvated Gα_{i1}^{myr} state, for instance, is the addition of two residues to the switch II region due to the downward shift of β2-β3 (Fig. 4.7). This elongation of switch II leads to a stability enhancement of this domain because of its ability to interact with the α3 helix (Fig. 4.3). The C-terminal region of the switch II region is structured via two or three helical turns that lead to a hydrophobic pocket for Trp211 (Fig. 4.7 and Fig. B.8). The obtained structure of the Gα_{i1}^{myr}:GTP switch II region is consistent with experimental data [Preininger et al., 2012] that show that Trp211 becomes located in a more hydrophobic environment when a myristoyl group is attached to the N-terminus of Gα_{i1}. Figure 4.7 depicts the conformation of switch II in both Gα_{i1}^{non}:GTPγS and Gα_{i1}^{myr}:GTP. From a structural inspection it can be concluded that Trp211 in Gα_{i1}^{myr}:GTP is mainly shielded from water by the switch II region itself and by α3-β5. However, due to the conformation of switch II in the Gα_{i1}^{non}:GTPγS X-ray structure, Trp211 appears to be solvent exposed, which is not in line with the hydrogen/deuterium exchange analysis performed by Preininger *et al.* [Preininger et al., 2012].

Besides the structural changes of the *Ras* domain, the AH domain also adjusts upon myristoyl

Table 4.1: Comparison of different Gα_{i1} states.

(I) Structures of Gα_{i1} in different states specified via their PDB code. The middle column describes the presented model. (a1) Residues 200 to 219 are shown in the depiction of the switch II region together with ligands: Mg^{2+} and GTP, GTPγS or GDP. (a2) Shows the sequence of residues that are included in the helical parts of the switch II region. (b) Shows if the structure is able to interact with the down stream protein adenylyl cyclase type 5. (c) Specifies which guanine nucleotide is interacting with the Gα_{i1} subunit. (d) Shows the location of the myristoyl group (MYR) in comparison to the rest of the protein. If MYR is in, the N-terminal tail is interacting with the hydrophobic binding pocket. If MYR is out, the myristoyl group is not present in the system and therefore the moiety is not interacting with the rest of the Gα_{i1} subunit.

	Structures^I		
Features	1GP2[Wall et al., 1995]	Model	1GIA[Coleman et al., 1994]
Switch II^{a1}			
Switch II^{a2}	207-216	208-215	204-214
State^b	Inactive	Active	Inactive
Ligand^c	GDP	GTP	GTPγS
MYR^d	Out	In	Out

binding. Especially the region that is not part of the interaction site with the *Ras* domain is significantly altered compared to the initial conformation. The regions in which the largest change can be observed are in the αA and the αB helices (Fig. 4.7c and Fig. B.5a). αB's initial conformation consists of four helical turns from Asp97 to Ala111 and two helical turns from Gly112 to Gly117. The final conformation of αB includes a double helical turn from Asp97 to Ala104 and three helical turns from Arg105 to Ala113. Since the secondary structure around the *Ras* domain and the AH domain is marginally changed, the only driving force behind these alterations in αB is the conformational change of the switch I region. During the classical MD simulation, the switch I region optimises its interactions with the environment, which leads to an altered conformation that is able to get into close proximity of the AH domain. Residues Glu115 and Glu116 in helix αB are biased towards the highly charged region around the GTP binding site that is located more on the *Ras* domain in the Gα_{i1}^{myr}:GTP complex than in the Gα_{i1}^{non}:GTP structure (Fig. 4.8). Furthermore, the switch I region is able to come closer to αB because of this conformational change. These two characteristics lead to a kinked helix that includes an unstructured region between Ala114 and Met119 that connects the αB helix to αC. This unstructured region contains Glu115 and Glu116 that can interact with switch I through interactions with Thr181 and the backbone of Lys180. Due to its structural changes, the αB helix modifies the surface and electrostatics of the protein and is able to interact with the switch I region.

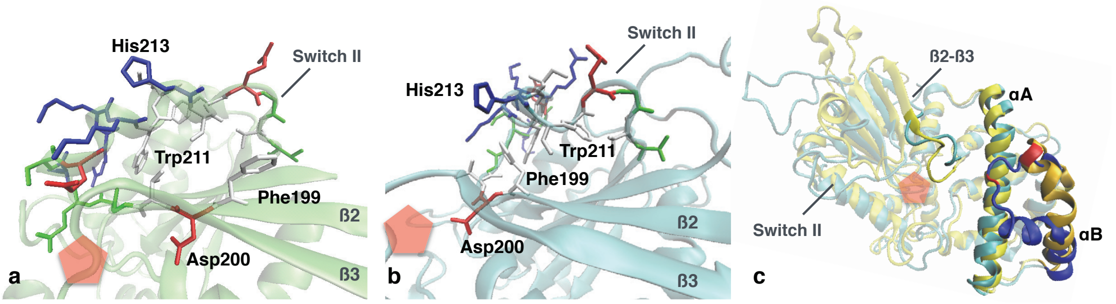


Figure 4.7: Comparison of the switch II region of Gα_{i1}^{myr}:GTP and Gα_{i1}^{non}:GTPγS together with the change conformation of the AH domain. (a) Detailed description of the switch II region of the 1GIA PDB structure. Non-polar residues are depicted in *white*, polar residues in *green*, positively charged residues in *blue* and negatively charged residues in *red*. The nucleotide binding site is shown via the *red* pentagon. (b) Detailed description of the switch II region of the solvated Gα_{i1}^{myr}:GTP structure. The color scheme for the residues and nucleotide interacting site is the same as in the *a* image. (c) Conformational changes in the AH domain when comparing the initial conformation (*yellow*) to a conformation of the protein after ~ 1.8 μs (*blue*). The *red* regions on the helices depict the location of Glu115 and Glu116. The *red* pentagon represents the location of the GTP binding site.

4.4.5 Electrostatic Properties and Protein-Protein Interaction Site

Besides the structural changes, also the electrostatic properties around the switch II region are significantly different when myristoylated and non-myristoylated active Gα_{i1} states are compared (Fig. 4.8). First, the positively charged region around the GTP binding site is more narrowly localised on the *Ras* domain in the soluble Gα_{i1}^{myr} conformation compared to the Gα_{i1}^{non}:GTPγS structure. Second, the area around the α3 helix is less negatively charged in the Gα_{i1}^{myr}:GTP complex with respect to the Gα_{i1}^{non}:GTPγS conformation. Third, the negatively charged patch created by the N-terminus near the C-terminal side of switch II (Fig. 4.8e-4.8h), is not present in Gα_{i1}^{non}:GTPγS as this region is resolved in the X-ray structure.

Commonly, Gα_{i1}'s switch II region, close to the GTP binding site, is proposed to be the main interaction site for down stream proteins like adenylyl cyclase. In the case of AC, two pseudo symmetrical sites are present on the protein of which one, the one located on the C2 domain, is known to interact with the activated form of Gα_s via its switch II region (Fig. B.10). Gel-filtration and activity experiments have shown that the interaction of AC5 with both G proteins is mainly non competitive and that Gα_s can only form a complex with C2 while Gα_{i1}^{myr} complexes solely with the C1 domain (Fig. B.11) [Dessauer et al., 1998; Preininger et al., 2012] . Hence, the C1 domain is proposed to interact with Gα_{i1}^{myr}:GTP's switch II region as X-ray structures of both activated Gα_s and Gα_{i1}^{non} are structurally similar (Fig. 4.8a, 4.8e). However, upon solvation, the conformation of Gα_{i1}^{myr} significantly changes compared to the non-myristoylated X-ray structure, including the switch II region. Therefore, a similar interaction site on Gα_{i1}^{myr}:GTP for C1 as the one on Gα_s:GTP for the C2 domain, located between α3 and switch II, appears to become less favourable as both switch II and the N-terminus are shielding this region (Fig. B.11 and Fig. 4.8c, 4.8g). This change in conformation could indicate that a different interaction site for AC type 5 exists on the soluble Gα_{i1}^{myr}:GTP structure. Such a view is also supported by the fact that AC type 5 is only inhibited by Gα_{i1}^{myr}:GTP (Table 4.1), as Gα_{i1}^{non}:GTP loses its inhibiting function and is unable to interact with AC5 [Taussig et al., 1993]. Furthermore, Gα_{i1}^{myr}:GTP is even able to inhibit AC5 when only the catalytic domains (C1:C2) are taken into account, which suggests that Gα_{i1}^{myr}:GTP is able to interact with AC5 in the cytosol [Dessauer et al., 1998]. Therefore, we propose here that the soluble form of Gα_{i1}^{myr}:GTP could be a good candidate for AC5 binding.

4.5 Conclusions

In the absence of myristoylation of the N-terminus, experimental studies have been able to elucidate the structure of active and inactive Gα_{i1}^{non}. However, this N-terminal modification is present under physiological conditions and is known to affect the subunit's conformation and function in the protein's GTP-bound state. In the case of inactive G$_{i1}^{myr}$:GDP, the complex is membrane bound, the myristoyl group interacts with the membrane and is not located

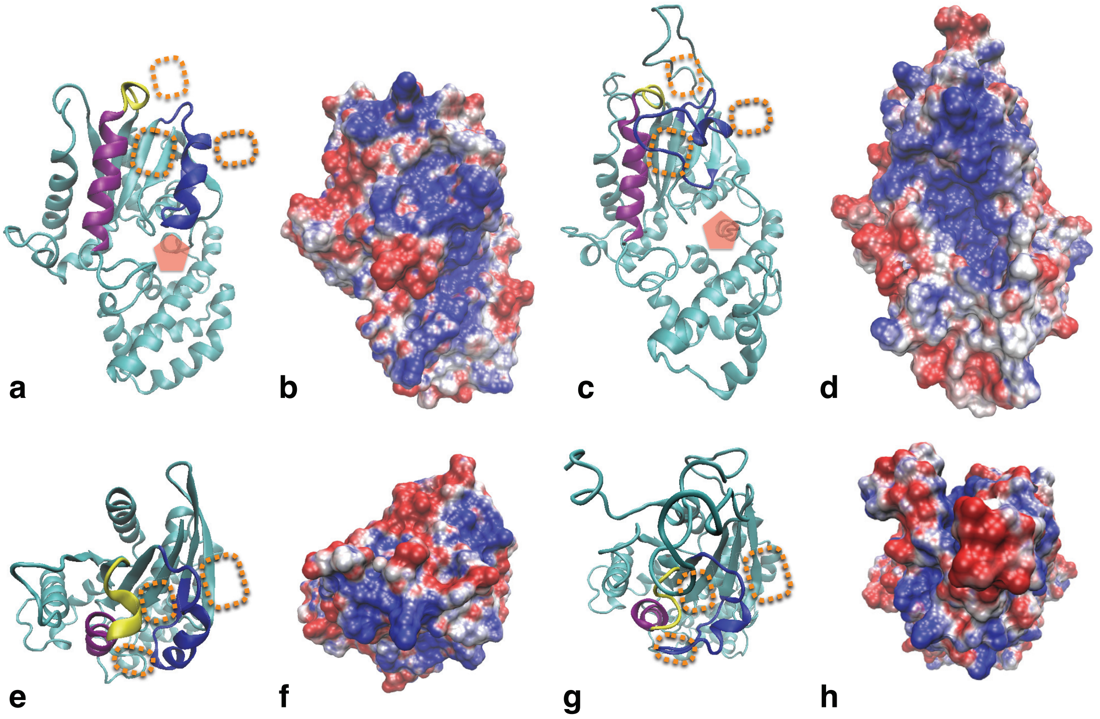


Figure 4.8: Comparison of the Gα_{i1}^{non}:GTPγS complex and Gα_{i1}^{myr}:GTP. Front (a) and top (e) view of the Gα_{i1}^{non} X-ray structure (PDB code 1GIA) in complex with GTPγS, a GTP analog. The pentagon indicates the location of the GTPγS molecule and the *orange* dotted rectangles represent the regions at which AC5 binds to Gα_s in the X-ray structure (PDB code 1CJK). The *blue* region shows the switch II region, the *purple* helix is α3 and the *yellow* part of the protein is α3-β5. (b, f) Electrostatic surface maps the front (b) and top (f) view of a Gα_{i1}^{non} X-ray structure (PDB code 1GIA) interacting with GTPγS. Three colours are employed: *blue* for positive values of the electrostatic potential, *white* for uncharged regions and *red* for negatively charged areas. The surface of the Gα_{i1} protein is contoured from -3 (*red*) to +3 (*blue*) kT/e based on the electrostatic potential present at the solvent accessible surface. (c, g) Front (c) and top (g) view of the Gα_{i1}^{myr} structure after about 1.8 μs. The colouring scheme is the same as in image *a*. (d, h) Electrostatic surface map of the front (d) and top (h) view of the Gα_{i1}^{myr} structure after about 1.8 μs of classical MD. The colour coding is the same as in image *b*.

close to the *Ras* domain of the Gα_{i1} subunit. However, the conformation of an active myristoylated α subunit can be affected by the myristoyl group as the α subunit can relocate to the cytosol due to the dissociation from the $\beta\gamma$ dimer and the depalmitoylation of the N-terminus, therefore providing the opportunity for the myristoyl group to interact with the α subunit. The performed 2 μs simulation of the conformation of active Gα_{i1}^{myr} shows that a solvated Gα_{i1}:GTP's structure is able to adopt a conformation in which the myristoyl moiety is embedded in a hydrophobic binding pocket on the protein itself. The presented conformation of the Gα_{i1}^{myr}:GTP complex provides a possible explanation for the disagreement between the Gα_{i1}^{non}:GTPγS X-ray structure and other experiments on active Gα_{i1}^{myr}, such as fluorescence and hydrogen/deuterium exchange experiments.

The structural rearrangements in solvated Gα_{i1}^{myr}:GTP present a new view on Gα_{i1}'s possibilities for protein-protein interactions as for its interaction with AC5, for instance. AC5 is present in high concentrations in the brain and heart, which has been associated to congestive heart failure and pain perception [Pierre et al., 2009; Sadana and Dessauer, 2008]. Inhibitors of AC5 are therefore proposed as possible drug target for treatment of heart failure and analgesics [Pierre et al., 2009]. Hence, obtaining a better view on how these proteins interact could be important for the development of AC5 inhibitors. Additionally, the presence of two active Gα_{i1} conformations, membrane-bound and solvated, constitutes a significant difference between Gα_{i1} and Gα_s as it is suggested that Gα_s only adopts one conformation of the *Ras* and the alpha helical domain with a structured or unstructured N-terminus in the active state, depending on its location in the cell. Consequently, although Gα_s is proposed to lose its structure of the N-terminus upon activation and depalmitoylation, the overall conformation of the protein could remain the same in the solvated state compared to the membrane-bound conformation as Gα_s, without its N-terminus, is still able to interact with AC5 [Tesmer et al., 1999]. This striking difference between Gα_{i1} and Gα_s could provide an alternative view on the function of G proteins in their active conformation. Moreover this study shows that lipid interactions can be crucial for the conformation of proteins and that through lipidation the function of structurally highly similar proteins can be altered and regulated.

5 Exploring the Inhibition Mechanism of Adenylyl Cyclase Type 5 by N-Terminal Myristoylated Gα_{i1}:GTP

Siri Camee van Keulen, Ursula Röthlisberger

5.1 Summary

Adenylyl cyclase (AC) is an important messenger involved in G-protein-coupled-receptor signal-transduction pathways. AC is regulated by stimulatory (Gα_s) and myristoylated inhibitory (Gα_i) G proteins. Understanding the interaction between Gα_{i1} and AC type 5 is important as it provides more insight into the unknown mechanism of AC5 inhibition, which is a possible target for the treatment of heart failure and analgesics. This study demonstrates via classical molecular dynamics simulations that the myristoylated Gα_{i1} structure is able to interact with adenylyl cyclase type 5 in a way that causes inhibition of the catalytic function of the enzyme.

5.2 Introduction

Many proteins are involved in cell communication of which one type is the G-protein-coupled receptor (GPCR), embedded in the membrane. GPCRs are part of a major signalling pathway, the GPCR signal-transduction pathway, which enables the transfer of a signal from the extracellular region to the intracellular side and is a key target for drug development. A large diversity of GPCRs can be found in nature as about 800 human genes are involved in storing different types of GPCRs that can interact with neurotransmitters, hormones or exogenous ligands, for example [Oldham and Hamm, 2008].

In the cytosol, G proteins are the first interaction partner of the membrane-bound activated GPCRs. G proteins are composed of an α, β and γ subunit. When a heterotrimeric G protein is activated by a GPCR, the trimer dissociates, resulting in an α subunit and a $\beta\gamma$ dimer [Galés et al., 2006]. Activated Gα subunits transport the signal from the membrane to other regions of the cell by stimulating or inhibiting reactions via protein-protein interactions. Besides direct activation by GPCRs, the function of G proteins can also be influenced by other environmental factors, such as lipidation. Permanent N-myristoylation, for instance, is known to change the structure and function of the inhibitory G-protein subunit Gα_{i1} in its active GTP-bound state [Degtyarev et al., 1994; Preininger et al., 2003; Dessauer et al., 1998; Taussig et al., 1993].

While a wide range of GPCRs exists, a relatively low diversity is present in the G protein family, e.g. the human body. The human body includes only a relatively small variety of 21 α, 6 β and 12 γ subunits [Oldham and Hamm, 2008]. The Gα subunits are divided into four major subfamilies based on their sequence homology and function [Simon et al., 1991]: stimulatory Gα_s, inhibitory Gα_i, Gα_q and Gα_{12}[Chen and Manning, 2001; Cabrera-Vera et al., 2003]. Overall the structures of the Gα subfamilies are similar (Fig. C.1), including a *Ras* domain and an alpha helical (AH) domain. The *Ras* domain is present in all members of the G-protein superfamily and can perform hydrolysis of GTP to GDP during deactivation of the Gα subunit, for example [Johnston and Siderovski, 2007]. In addition, the domain includes an interaction site for GPCRs as well as regions that can interact with the $\beta\gamma$ dimer. Moreover, the *Ras* domain can also undergo lipidation. Except for Gα_t, all Gα proteins are able to reversibly bind a palmitate to their N-terminal helix. Besides palmitoylation, Gα_i can also permanently bind a myristoyl moiety to the N-terminus that appears to be crucial for the function of the subunit [Cabrera-Vera et al., 2003; Preininger et al., 2003, 2012; Dessauer et al., 1998].

The AH domain is unique to the Gα subfamilies, which is composed of six α helices and interacts with the *Ras* domain when GTP or GDP is present. However, this interaction between the AH and *Ras* domain is weakened when a nucleotide is absent in Gα's active site [Dror et al., 2015; Louet et al., 2012; Schröter et al., 2015]. The high structural similarity among members of the Gα subfamilies is illustrated by aligning the X-ray structures of stimulatory Gα_s and inhibitory Gα_{i1}, resulting in a root mean square deviation (RMSD) of only 1.07 Å between the

Cα atoms of the two structures (Fig. C.1) [Kaya et al., 2014; Sunahara et al., 1997b]. Hence, from a comparison of the structures it is difficult to conclude what the origin is of their inverse action, i.e., how the structure can be related to a stimulatory respectively an inhibitory effect.

An example of a protein in which both Gα_i and Gα_s are important for regulation is adenylyl cyclase (AC). Ten isoforms of AC are known of which nine are membrane-bound (AC1-9) and one is soluble (sAC) [Sunahara and Taussig, 2002]. These different types of AC are found throughout the body in different concentrations. AC5, for instance, is present in high quantities in the brain, the spinal cord and the heart, and is associated to congestive heart failure and pain perception [Pierre et al., 2009; Sadana and Dessauer, 2008]. G proteins have the ability to either stimulate (G$_s$) or inhibit (G$_i$) adenylyl cyclases' conversion of adenosine triphosphate (ATP) to cyclic adenosine monophosphate (cAMP) and pyrophosphate [Robinson et al., 1968].

ACs consist of two membrane-bound regions, each built from six trans-membrane domains, and a catalytic region in the cytosol that includes two pseudo-symmetric domains, C1 and C2 [Hurley, 1999]. GTP-bound Gα_s is known to bind to the C2 domain for which the interaction site is known from X-ray structures of Gα_s interacting with AC [Tesmer et al., 1999]. Such data is absent for the case of Gα_i. In the absence of direct experimental information, a putative interaction site of GTP-bound Gα_i has been suggested in analogy to the known structure of the complex of Gα_s and AC (Gα_s:AC) as the pseudo-symmetric site on the C1 domain. However, how the interaction of Gα_i on the C1 domain should induce inhibition is not obvious (Fig. C.2) [Dessauer et al., 1998]. Furthermore, since with this hypothesis the interaction sites of Gα_{i1} and Gα_s are highly similar in addition to their structures, it is unclear how the α subunits can differentiate the two binding sites on AC and what the cause is of the stimulatory versus the inhibitory effect induced by the subunits.

A factor that could play an important role in differentiating the action of Gα_i and Gα_s is the difference in lipidation of both subunits. Although the X-ray structures in the Protein Data Base (PDB) [Berman et al., 2000] of the active inhibitory and stimulatory Gα subunits tightly align, the N-terminus, which is not resolved for Gα_i or Gα_s, is not myristoylated during the expression process of Gα_i as lipidation can hinder crystallisation [Preininger et al., 2003]. Hence, it is not clear to what extent the missing N-terminal myristoyl moiety affects the Gα_i structure of the remaining protein while the bound myristoyl group has been known to be crucial for Gα_i's conformation and function as the ability to interact with AC5 is abolished upon removal of the myristate [Dessauer et al., 1998; Preininger et al., 2003; Taussig et al., 1993]. Classical molecular dynamics (MD) simulations of myristoylated Gα_{i1}:GTP (Gα_{i1}^{myr}) show that myristoylation can have a significant effect on the conformation of the subunit [van Keulen and Rothlisberger, 2016]. The findings suggest the possibility of an alternative novel interaction mode and open up new possibilities for selective interactions with AC. This is because the found structural changes in the classical MD simulations of active Gα_{i1}^{myr} [van

Keulen and Rothlisberger, 2016] suggest that the subunit will not be able to interact with C1 in a similar way as Gα_s interacts with C2. In addition, MD simulations of active non-myristoylated Gα_{i1} interacting with AC5's C1 domain suggest that this complex is highly unstable [Lavery].

Here, we investigate the interaction between Gα_{i1}^{myr} and AC, using classical MD simulations. To this end, the initial structure of Gα_{i1}^{myr} was taken from reference van Keulen and Rothlisberger [2016]. Gα_{i1}^{myr} can inhibit only particular isoforms of AC: AC1, AC5, AC6 [Sunahara et al., 1996]. In this study AC5 is used because X-ray structures of AC's catalytic domains are composed of isoforms AC2 and AC5. Ca. 16 AC structures can be found in the PDB with different resolutions and/or crystallisation conditions. All available structures have been co-crystallised with a Gα_s subunit and correspond therefore to activated conformations at various levels of activation, depending on the nature of bound cofactors (e.g. cofactor-free complex of AC, substrate-bound AC complex).

When AC5 becomes active, roughly three conformational options are possible: a complex of Gα_s and AC5 (Gα_s:AC5), Gα_s in complex with AC5 with bound ATP (Gα_s:AC5:ATP), or a complex of Gα_s and AC5 bound to the reaction products cAMP and pyrophosphate (Gα_s:AC5:cAMP). Currently, it is not known which one of these forms is most likely to interact with Gα_{i1}^{myr}, or if Gα_{i1}^{myr} can inhibit all of them. In this study, the structure of the AC5 protein was taken from a crystal structure of the cofactor-free Gα_s:AC5 complex. This AC5 conformation was used as it can be viewed as the first and the least-stable active conformation among the three forms, which increases the possibility of observing conformational changes on the time scales that can be reached via classical MD. The AC5 structure was employed to build a Gα_{i1}^{myr}:AC5 complex and to explore if the binding of Gα_{i1}^{myr} is able to reverse the active conformation initially induced by Gα_s. In order to verify which changes are due to the interaction of AC5 with Gα_{i1}^{myr} and which alterations are a result of the removal of Gα_s, a second simulation of AC5, with the Gα_s subunit removed, was performed on the same time scale as the Gα_{i1}^{myr}:AC5 complex.

Hence, in this study the impact of the presence of myristoylated Gα_i on the function of AC5 is explored via investigating the conformational features of the Gα_{i1}^{myr}:AC5 and the free AC5 complex (a system that only includes AC's catalytic region in solution) in comparison with the Gα_s:AC structure. This Gα_{i1}^{myr}:AC5 complex has been obtained via docking the Gα_{i1}^{myr} structure on to the C1 domain of AC5. Already the initial docking results confirm the possible importance of the myristoyl-induced structural changes of Gα_{i1}^{myr} as a new interaction mode for Gα_{i1}^{myr} could be identified. The comparison of the performed classical MD simulation (2.5 μs) of the Gα_{i1}^{myr}:AC5 complex and the free AC5 system suggest two possible ways of AC inhibition. First, active Gα_{i1}^{myr} seems to inhibit AC's conversion of ATP to cAMP by preventing active-site formation as the Gα_{i1} subunit perturbs the conformation of the active site on the C1/C2 interface. Second, the effect of Gα_{i1}^{myr} on the AC structure leads to a closed conformation of the Gα_s binding site on C2, decreasing the probability of Gα_s association and thus of a counter-balancing re-stimulation of the AC5 activity. Taken together, the observed events lead to a suggestion for a putative Gα_{i1}^{myr} inhibition mechanism of AC5 (Fig. 5.7) in which lipidation is crucial for Gα_{i1}^{myr}'s function and its protein-protein interactions. This gives a possible indication that lipidation could play a significant role in regulating G protein function and thus signalling in G protein mediated pathways [Dessauer et al., 1998; Preininger et al., 2003; Taussig et al., 1993].

5.3 Methods

5.3.1 Initial Structures

The PDB structure 1AZS, composed of the catalytic domains of the ATP-free Gα_s:AC complex, was used as a template, including 1AZS's C1 and C2 domain, for the initial AC5 structure of the Rattus norvegicus (UniprotKB Q04400) [Tesmer et al., 1999; Eswar et al., 2006; Martí-Renom et al., 2000]. The missing loop in the C2 domain was modelled using a snippet of the PDB structure 1J8M [Montoya et al., 2000]. The structure of the Rattus norvegicus Gα_{i1}^{myr} subunit (UniprotKB P10824) interacting with GTP and Mg^{2+} was taken from reference van Keulen and Rothlisberger [2016] (Fig. C.3 and Fig. C.4).

5.3.2 Docking of Gα_{i1} with AC5

The HADDOCK web server [De Vries et al., 2010] was used for docking ten conformations of the myristoylated Gα_{i1}:GTP complex with the catalytic domains of AC5 of Rattus norvegicus. The active region of Gα_{i1} was defined in HADDOCK as a large part of the AH domain (112-167), the switch I domain (175-189) and the switch II domain (200-220), allowing for a large unbiased region on the Gα_{i1}^{myr} protein surface to be taken into account during docking. The active region of AC5's C1 domain was defined as the α1 helix (479-490) and the C-terminal region of the α3 helix (554-561) because experimentally it has been found that Gα_{i1}^{myr}:GTP is unable to interact with C2 and its main interactions with AC are with the C1 domain [Dessauer et al., 1998]. Passive residues, residues that could take part in protein-protein interaction, were defined as residues around the active residues that are on the protein surface and within a radius of 6.5 Å of any active residue [De Vries et al., 2010]. Ten Gα_{i1}^{myr} snapshots were used for docking with a time interval between each snapshot of 0.5 ns. These snapshots were taken at the end of the Gα_{i1}^{myr} classical MD trajectory (around 1.9 μs) discussed in reference van Keulen and Rothlisberger [2016].

The initial Gα_{i1}^{myr}:AC5 complex for the classical MD simulations was selected based on (1) the absence of overlap between the C2 domain and Gα_{i1}^{myr}:GTP, (2) no overlap with the GTP binding region and the interaction with C1 and (3) presence of similar complexes in the top-ten docking results of the docking calculations performed for all ten Gα_{i1}^{myr}:GTP snapshots. The first property of the selection criteria is important since Gα_{i1}^{myr}:GTP is unable interact with C2 [Dessauer et al., 1998]. The second criterium has been defined since GTP is located in the active site of Gα_{i1}^{myr} in the classical MD simulations, but was not incorporated in the docking procedure because this is not possible in HADDOCK. Therefore, no overlap between the GTP binding site and the C1 domain should be present in the docking result as otherwise the GTP molecule will not be able to fit in Gα_{i1}^{myr}'s active site. The last criterium is the presence of similar Gα_{i1}^{myr}:AC complexes of the selected complex in all top-ten docking results which increases the probability that complexation of the two proteins is not conformation specific, but is robust as similar complexes can be obtained using different conformations of Gα_{i1}^{myr}.

5.3.3 Classical Molecular Dynamics Simulations

The Gα_{i1}^{myr}:AC5 complex together with the two Mg^{2+} ions and GTP was used to simulate the protein complex for 2.5 μs at 310 K and 1 bar using a Nosé-Hoover thermostat and an isotropic Parrinello-Rahman barostat. Additionally, about 68 000 water molecules and 150 mM KCl are present in the simulated system. The force fields used for the protein and the water molecules are AMBER99SB [Hornak et al., 2006] and TIP3P [Jorgensen et al., 1983], which were employed by Gromacs 4.6.6 [Bekker et al., 1993; Hess et al., 2008] to perform the runs. For GTP, the force field generated by Meagher *et al.* was used [Meagher et al., 2003]. The adjusted force field parameters for K^+ and Cl^- were taken from Joung *et al.* [Joung and Cheatham III, 2008]. The Mg^{2+} parameters originated from Allnér *et al.* [Allnér et al., 2012] and the parameter set for the myristoyl group was taken from reference van Keulen and Rothlisberger [2016].

Electrostatic interactions were calculated with the Ewald particle mesh method with a real space cutoff of 12 Å. Bonds involving hydrogen atoms were constrained using the LINCS algorithm [Hess, 2008]. The time integration step was set to 2 fs.

The free AC5 system was also simulated with the same setup as the Gα_{i1}^{myr}:AC5 complex. The system was solvated in 30 000 water molecules and 150 mM KCl concentration. The initial location of Mg^{2+} in the active site of the enzyme was the same as in the Gα_{i1}^{myr}:AC5 complex system.

5.3.4 Structure Superpositions and Images

Multiprot [Shatsky et al., 2004] and VMD [Humphrey et al., 1996] were used to align protein structures. Uniprot [Consortium, 2014] was used to align protein sequences. Images were prepared with VMD [Humphrey et al., 1996].

5.4 Results and Discussion

5.4.1 Classical Molecular Dynamics Simulations of Gα_{i1}^{myr}:AC5 Complex

Gα_{i1}^{myr}'s proposed interaction site in the initial Gα_{i1}^{myr}:AC5 structure (see *Supporting Results and Discussion* Section in Appendix C) appears to affect the conformation of C1 in a different way than Gα_s stabilises the C2 domain (Fig. C.2 and Fig. C.4). Unlike Gα_s, Gα_{i1}^{myr} is not located between the helices of AC5's catalytic domain, but appears to clamp the C1 domain into its inactive conformation. Gα_{i1}^{myr} is positioned around AC5's α3, interacting with α1, α2, and α3 via its switch I, II and III domain together with the C-terminal domain of αB (Fig. C.4). Since C1's α1 helix appears to decrease its distance with respect to the C2 domain when an ATP analog is present in the active site (Fig. C.3c, d), the interactions between C1 and Gα_{i1}^{myr} in the initial Gα_{i1}^{myr}:AC5 complex could suggest that Gα_{i1}^{myr} is able to inhibit ATP's conversion by preventing AC5's α1 to rearrange upon ATP binding due to Gα_{i1}^{myr}'s interaction with C1's α1.

The stability of the docked complex and the effect of Gα_{i1}^{myr} association was verified via investigating the conformational changes in the complex through classical MD simulations on the μs time scale. The X-ray structure used for the initial conformation of AC5, was co-crystallised with Gα_s but is not interacting with substrate or products from the ATP conversion reaction (see Section 5.2). Consequently, this conformation of AC5 could be viewed as a structure that is present before ATP association, but is already in an active form, due to its interactions with Gα_s. Because of the crystallisation circumstances used to obtain the selected AC5 structure, the found conformational changes in the catalytic region of AC5 in the Gα_{i1}^{myr}:AC5 complex are compared to the conformation found in the classical MD trajectory of the free AC5 system and the Gα_s:AC X-ray structure in order to verify which structural alterations are due to the presence of Gα_{i1}^{myr} and which changes are on the other hand the result of the absence of Gα_s.

In order to closer mimic a protein system with which ATP or a product such as pyrophosphate would be able to interact, a Mg^{2+} ion was added to the active site of the Gα_{i1}^{myr}:AC5 and the free AC5 system. Throughout the MD simulations of free AC5 and Gα_{i1}^{myr}:AC5, the Mg^{2+} ion remained stable in the active site of AC5. Results on the stability of this Mg^{2+} ion in AC5's active site can be found in the *Supporting Results and Discussion* Section in Appendix C.

Protein-Protein Interface of the Gα_{i1}^{myr}:AC5 Complex

The Initial Gα_{i1}^{myr} Conformation is Very Stable over the Entire Course of the Classical MD Trajectory Since the starting structure of the Gα_{i1}^{myr}:AC5 complex is unrelaxed, the conformation of the Gα_{i1}^{myr} subunit and the Gα_{i1}^{myr}/AC5 interface have been investigated during the MD trajectory to study the stability of the Gα_{i1}^{myr}:AC5 complex and the effect of the AC5 interaction on the Gα_{i1}^{myr} structure. In fact, the structure of Gα_{i1}^{myr} only changes minimally by a slight adjustment in the orientation of the alpha helical domain (Fig. 5.1b and Fig. C.7). A striking feature of the interface between Gα_{i1}^{myr} and C1 is the fact that the switch regions, I, II and III, remain involved in AC5 binding, as well as the region on the alpha helical domain that rearranged upon myristoyl binding (Fig. 5.1a, b) [van Keulen and Rothlisberger, 2016]. Gα_{i1}^{myr} is stabilised on AC5 via the C1 domain without major interactions with C2 (Fig. 5.1c). The relative orientation of Gα_{i1}^{myr} with respect to the C1 domain stabilises after ~ 400 ns (Fig. C.7). This slight orientational repositioning is probably a consequence of the relocation of C2's β7-β8 loop, which could be due to the removal of the Gα_s subunit from the C2 domain in the initial complex because this β7-β8 loop displacement is present in Gα_{i1}^{myr}:AC5 as well as free

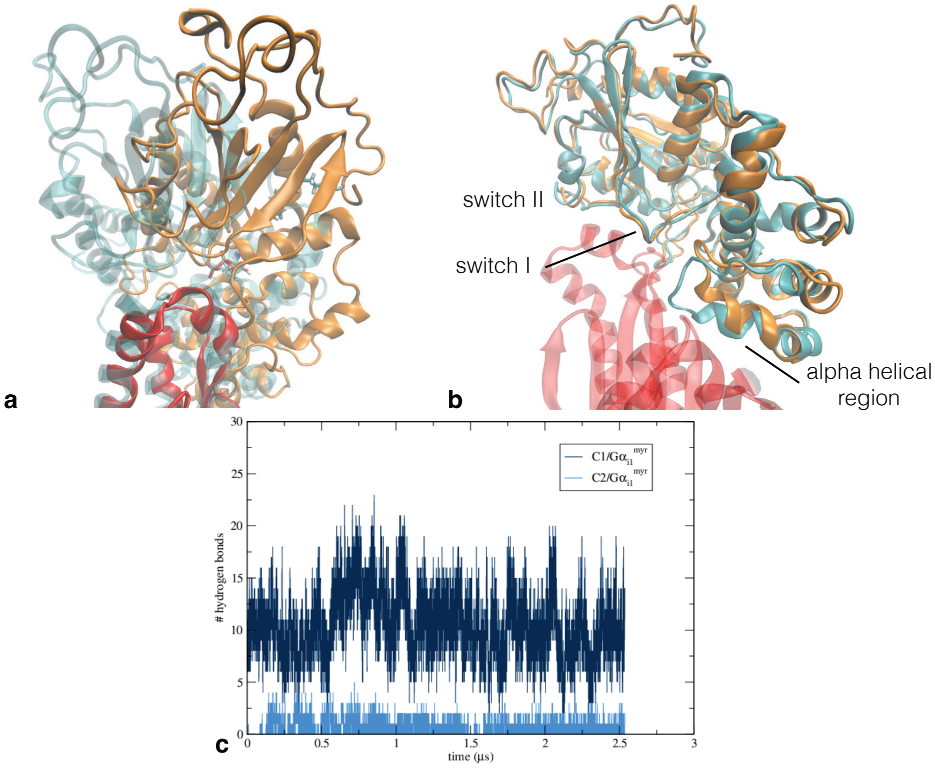


Figure 5.1: Gα_{i1}^{myr}'s interactions with AC5 are stable over the course of the classical MD trajectory. (a) Aligned structures of the docked model (cyan) and the Gα_{i1}^{myr}:AC5 complex after ~ 1.7 μs (orange and red). The structures were aligned on the C1 domain, residues 456 to 644, as this domain's RMSD is low over the course of the simulation (Fig. 5.3). (b) Aligned structures of the docked model (cyan) and the Gα_{i1}^{myr}:AC5 complex after ~ 1.7 μs (orange and red). The structures were aligned on the Gα_{i1}^{myr} subunit (residues 34 to 334) in order to show the change in the conformation of Gα_{i1}^{myr}. (c) Number of hydrogen bonds between Gα_{i1}^{myr} and C1 and Gα_{i1}^{myr} and C2 that are present during the classical MD trajectory.

AC5 (Fig. 5.2).

Besides the orientation of Gα_{i1}^{myr}, also minor alterations can be observed in Gα_{i1}^{myr}'s active site. The Mg^{2+} ion, near the GTP binding site (Fig. C.8), is moderately altered compared to the free Gα_{i1}^{myr} system [van Keulen and Rothlisberger, 2016]. In the Gα_{i1}^{myr}:AC5 both of Asp200's oxygens, OD1 and OD2, are able to interact with Mg^{2+}, while in free Gα_{i1}^{myr} only one of Asp200's oxygens is interacting with the Mg^{2+} ion [van Keulen and Rothlisberger, 2016]. The interaction between Mg^{2+} and Asp200's second oxygen in Gα_{i1}^{myr}:AC5, OD2, can be temporarily perturbed by a water molecule, which leads to sudden jumps in the Mg^{2+} and Asp200OD2 distance (Fig. C.8).

Structural Impact on AC5's Active Site

Gα_{i1}^{myr}'s Interactions with C1 Impact the Entrance of the ATP Binding Site The active site of AC5 is located on the interface between C1 and C2. Structural changes of both domains upon Gα_{i1}^{myr} binding can potentially influence the protein's activity. When comparing the root mean square deviation of the two domains with respect to AC'5 initial conformation, it is clear that in the free AC5 as well as in the Gα_{i1}^{myr}:AC5 complex the C2 domain experiences more changes than C1 (Fig. 5.3). In the X-ray structure, used as a template to construct the initial AC5 structure, the C2 domain is interacting with a Gα_s subunit. The removal of Gα_s from the C2 domain could affect the RMSD of the domain in the Gα_{i1}^{myr}:AC5 and free AC5 systems as

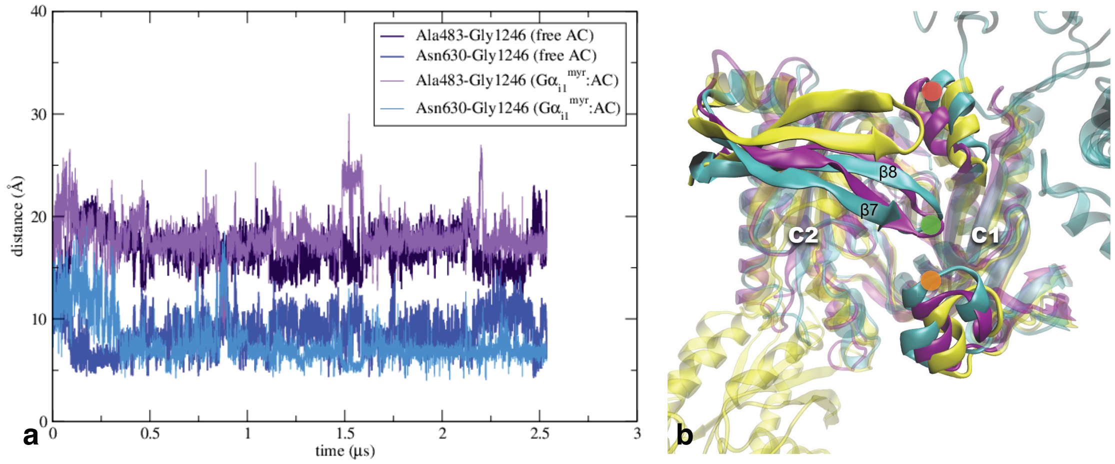


Figure 5.2: Change of location of C2's β7-β8 loop occurs in both Gα_{i1}^{myr}:AC5 and free AC5 systems. (a) Graph of the distances between the α carbon of Gly1246 (green dot in image *b*) and the α carbons of Ala483 (red dot in image *b*) and Asn630 (orange dot in image *b*). (b) β7-β8 loop's relocation in the free AC5 system (purple), the Gα_{i1}^{myr}:AC5 complex (cyan) and PDB structure 1AZS (yellow). The location of the residues used in image *a* are assigned according to the Gα_{i1}^{myr}:AC5 structure.

the initial structure of AC5 is influenced by the presence of Gα_s.

During classical MD the first alteration to C2's initial structure that can be observed in both Gα_{i1}^{myr}:AC5 and free AC5 is the relocation of the β7-β8 loop, positioned on the cytosolic side (Fig. 5.2). The location of C2's β7-β8 loop is important for the active conformation of the active site on the C1/C2 interface (Fig. 5.2 and Fig. C.3). Loop relocation and the accompanying movement of the two domains appear to have a deactivating effect on the active site (Fig. 5.2 and Fig. C.3b). For instance, a residue that is known to be part of the active site, Lys1245, located in the β7-β8 loop, is unable to maintain its orientation towards the ATP binding site (Fig. C.9). A reason for this conformational change in the Gα_{i1}^{myr}:AC5 and free AC5 simulations could be the absence of Gα_s at the C2 domain that destabilises β7-β8's location. The presence of Gα_{i1}^{myr} appears to increase the stability of the β7-β8 loop relocation compared to the free AC5 system (Fig. 5.2).

An alteration that also occurs around AC5's active site in the Gα_{i1}^{myr}:AC5 system is the decrease in distance between C2's α4 helix and C1's β2-β3 loop (Fig. 5.4, Fig. C.4). This change in the C1/C2 interface could make the positioning of ATP in the active site less favourable because ATP's adenine moiety, which is positioned between C2's α4 helix and C1's β2-β3 loop when ATP interacts with AC5 (Fig. C.3d), is unlikely to fit between the C1/C2 interface due to the diminished distance between the C1 and C2 regions compared to the initial AC5 structure. Although the free AC5 system also undergoes a change in this region, the α4 helix of C2 remains closer to the X-ray location than in the Gα_{i1}^{myr}:AC5 complex (Fig. 5.4).

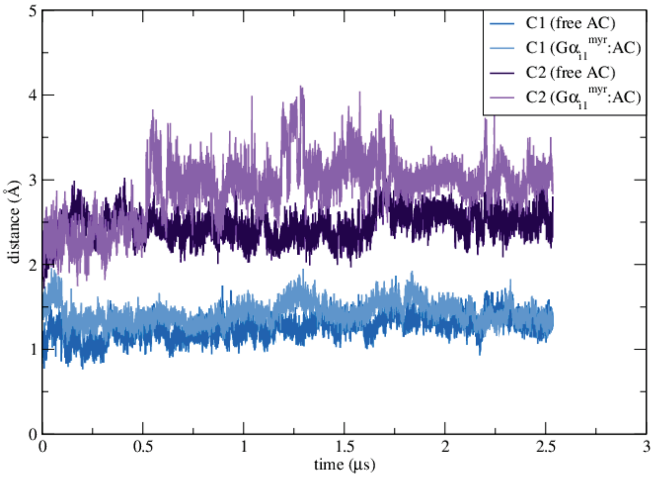


Figure 5.3: Significant difference between RMSD values of free AC5 and Gα_{i1}^{myr}:AC5 for the C2 domain. Root mean square deviations of the backbone of the C1 and the C2 domain. In the RMSD calculation the residues between 463 to 644 were taken into account for the C1 domain and the residues between 1065 to 1135 and 1145 to 1257 were used for the C2 domain.

In summary, the conformational changes at the C1/C2 interface of Gα_{i1}^{myr}:AC5 seem to affect AC5's active site by interfering with ATP binding. In fact, the new position of the β7-β8 loop can even block the active site entrance, which could prevent ATP entry. The relocation of C2's β7-β8 loop, present in the Gα_{i1}^{myr}:AC5 and the free AC5 systems, could be due to the removal of Gα_s, which interacts with the C2 domain in the template X-ray structure. The interaction between Gα_s and AC5 seems to stabilise the position of β7-β8 at the C1/C2 interface, close to C1's α1. As C2's β7-β8 is part of the active site, the position of β7-β8 could be viewed as a stimulation feature that can be switched on by Gα_s or switched off by the absence of Gα_s (free AC5) or by the presence of Gα_{i1}^{myr}, which stabilises the relocation of the β7-β8 loop even more (Fig. 5.2).

Gα_{i1}^{myr}'s Interactions with C1 Impact ATP's Binding Site around the Adenine Moiety of ATP

While the RMSD of free AC5 and Gα_{i1}^{myr}:AC5 compared to the initial AC5 structure (obtained via the Gα_s:AC X-ray structure) appear similar for C1, the C2 domain of Gα_{i1}^{myr}:AC5 diverges more from the initial structure than free AC5 (Fig. 5.3). One of the major differences between the C2 domain of the free AC5 system and the Gα_{i1}^{myr}:AC5 complex lies on the membrane side of the proteins. In the case of free AC5, C2's β4-β5 loop is interacting with the C1 domain,

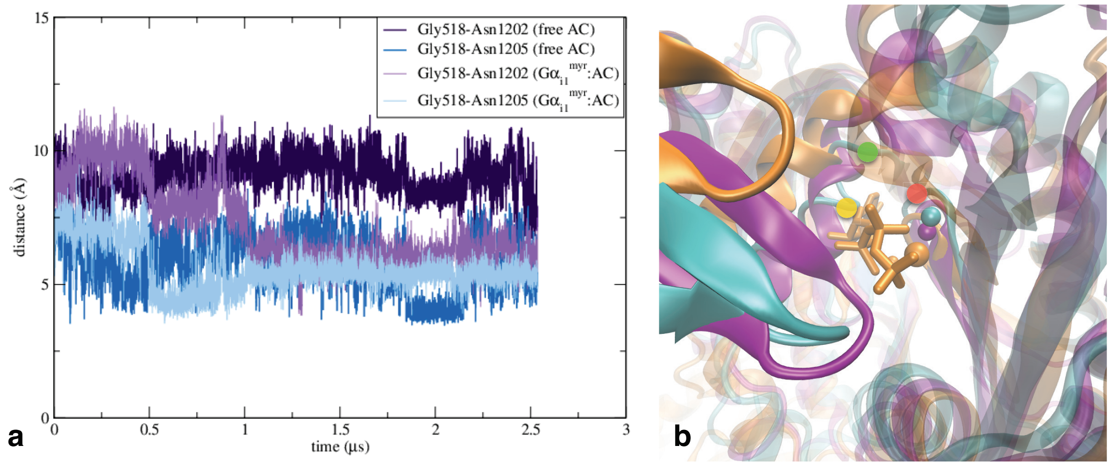


Figure 5.4: Rearrangements of AC5's active site differ between the Gα_{i1}^{myr}:AC5 complex and free AC5 system. (a) Graph of the distances between the α carbon of Gly518 (red dot in image *b*) and the α carbons of Asn1202 (green dot in image *b*) and Asn1205 (yellow dot in image *b*). The respective distances in PDB structure 1CJK of Gly518-Asn1202 (Gly439-Asn1022 in PDB 1CJK) and Gly518-Asn1205 (Gly439-Asn1025 in PDB 1CJK) are 11 Å and 8.5 Å. (b) Detail of AC's active site of the free AC5 system (purple), the Gα_{i1}^{myr}:AC5 complex (cyan) and PDB structure 1CJK (orange). The location of the residues used in image *a* are assigned according to the Gα_{i1}^{myr}:AC5 structure. In the active site the location of the Mg^{2+} are shown for the three structures as well as the position of the ATP analog from the fully activated AC5 structure (PDB code 1CJK).

while in the Gα_{i1}^{myr}:AC5 complex, the loop does no longer interact with the C1 domain, leading to an unfavourable ATP binding site at the C1/C2 interface (Fig. 5.5a, b, c). A weakening of the ATP binding site at the C1/C2 interface is also apparent from C1's interactions with the C2 domain near the active site region (Fig. C.9). Asp1198, for example, which is important for stabilising ATP's adenine moiety in the Gα_s:AC complex, reorients as the residue is part of the β4-β5 loop. Due to Asp1198's change in location, Lys1124, which is interacting with Asp1198, alters its orientation as well. Lys1124 also influences the stability of the active site in Gα_s:AC as the residue stabilises ATP in a similar fashion as Asp1198 (Fig. C.9).

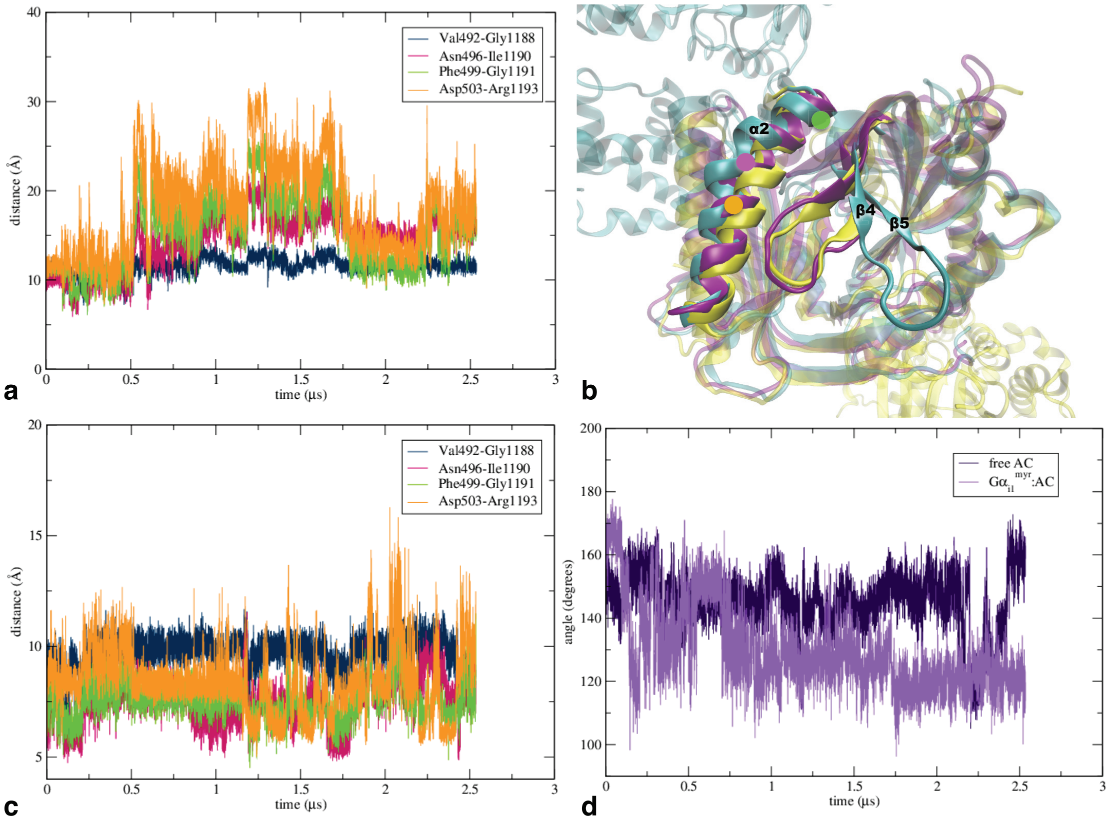


Figure 5.5: Conformational changes on membrane side of AC5 show C2's loop dissociation, which only occurs in Gα_{i1}^{myr}:AC5. (a) Graph of the distances between the α2 helix of C1 and the β4-β5 loop of C2 in the Gα_{i1}^{myr}:AC5 structure (see image *b*) (b) AC5's membrane side of the free AC5 system (purple), the Gα_{i1}^{myr}:AC5 complex (cyan) and PDB structure 1AZS (yellow) in which the α2 helix of C1 and the β4-β5 loop of C2 are highlighted. In the Gα_{i1}^{myr}:AC5 complex the location of the residues used in the angle calculation of image *d* are represented by a green (Ala488), pink (Leu495) and orange (Phe499) dot. (c) Graph of the distances between the α2 helix of C1 and the β4-β5 loop in the AC5 system (see image *b*). (d) Angle between three helical turns, including Cα atoms of Ala488, Leu495 and Phe499, in which the kinking of the α2 helix of C1 takes place (see image *b*).

Although C1's RMSD is low (Fig. 5.3 and Fig. C.7), a significant change in conformation can be observed close to the Gα_{i1}^{myr}/C1 interface where a kink in α2 occurs, which is less pronounced in free AC5 (Fig. 5.5d). However, overall, the C2 domain seems to be affected most by the absence of Gα_s and the presence of the inhibitory Gα. This observation is in line with the hypothesis that Gα_{i1}^{myr} is able to constrain C1's conformation via its tight interactions with this domain, leading to a perturbation and destabilisation of the active site on the C1/C2 interface (Fig. 5.5 and Fig. C.8). This change of the C1/C2 interface prevents the catalytic domains to sample the conformation in which ATP could be positioned in the active site due to a decrease in distance between C2's α4 helix and C1's β2-β3 loop and a relocation of C2's β4-β5 and β7-β8 loops (Fig. 5.2, Fig. 5.4 and Fig. 5.5), which play important roles in the construction of the active site.

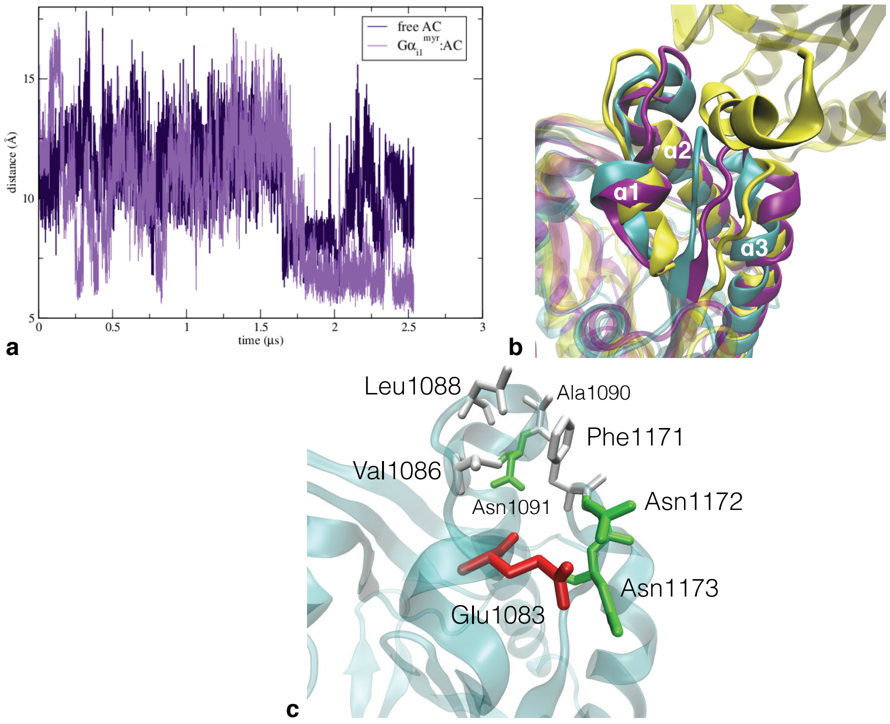


Figure 5.6: Conformational changes around the G$_s$ binding site on C2 show distinct events of closure in the Gα_{i1}^{myr}:AC5. (a) Graph of the distance between the α2 and the α3 helix of C2 including Cα atoms of Asn1091 and Phe1171, of free AC5 and Gα_{i1}^{myr}:AC5. A detailed representation of the G$_s$ binding site is shown in image *c*. (b) the G$_s$ binding site of the free AC5 system (purple), the Gα_{i1}^{myr}:AC5 complex (cyan) and PDB structure 1AZS (yellow) in which the G$_s$ subunit is also shown. (c) Detail of the G$_s$ binding site of Gα_{i1}^{myr}:AC5 in which residues that are involved in the closing of the binding site are shown.

Presence of Gα_{i1}^{myr} Induces Closure of Gα_s's Interaction Site on AC5's C2 Domain Besides the direct conformational changes in AC5's active site another mechanism that could induce inhibition is decreasing the probability of Gα_s binding to the C2 domain. In free AC5 as well as in Gα_{i1}^{myr}:AC5, the Gα_s site seems to become less favourable for Gα_s binding since the distance between α2 and α3, α2-α3, is significantly decreased with respect to the X-ray structure of the Gα_s:AC5 complex (Fig. 5.6a, b). In the free AC5 system α2-α3 changes from around 16 Å (initial distance between the Cα carbons of Asn221 and Phe1171) to an average distance of ~11 Å (Fig. 5.6a,b). In the Gα_{i1}^{myr}:AC5 simulations, the distance between the two residues can decrease even more severely to a distance of ~ 7 Å (Fig. 5.6a,b), leading to a closed Gα_s binding site conformation. Important residues that stabilise this closed conformation of the Gα_s binding site in Gα_{i1}^{myr}:AC5 are AC5's: Glu1083, Leu1088, Ala1090, Phe1171, Asn1172 and Asn1173 (Fig. 5.6d). Hence, due to the interaction with Gα_{i1}^{myr}, the catalytic domains of AC5 appear to deactivate AC5's catalytic ability via deformation of the active site as well as to sample closed Gα_s binding site conformations.

5.4.2 Possible Mechanism of Gα_{i1}^{myr} Inhibition

The simulations of Gα_{i1}^{myr}:AC5 in comparison with free AC5 and the Gα_s:AC X-ray structure demonstrate that the first step in decreasing AC5's activity is the relocation of the β7-β8 loop (Fig. 5.7, step one). In fact, the β7-β8 loop seems to have an important role for the stimulatory response since the presence of Gα_s leads to the stabilisation of the loop, forming ATP's binding site (Fig. 5.7, starting conformation of AC in left panel) [Tesmer et al., 1999]. This loop conformation is lost as soon as Gα_s is reduced as observed for both free AC5 and Gα_{i1}^{myr}:AC5. In step two of figure 5.7 the Gα_{i1}^{myr}:AC5 complex undergoes a rearrangement in the C2 domain (absent in free AC5), which leads to a further perturbation of AC5's active site and a destabilisation of the C1/C2 interface. The classical molecular dynamics simulations also show that in the presence of Gα_{i1}^{myr}, there appears to be a decrease in probability for Gα_s association (Fig. 5.7, step 3 and Fig. 5.6). Through these rearrangements Gα_{i1}^{myr} could deactivate AC5 as well as decrease the probability of reactivation via Gα_s.

5.5 Conclusions

The results of this study suggest that Gα_{i1}^{myr} deactivates adenylyl cyclase type 5 via constraining C1's active site region. Inhibition and stimulation of AC5 appear to follow different pathways. While Gα_s binds between the helices of C2, increasing the stability of the C1:C2 dimer, Gα_{i1}^{myr} is able to clamp the helices of the C1 domain, promoting an inactive conformation of AC5's catalytic domains and a possible decrease in affinity for Gα_s on the C2 domain. Structurally, Gα_s and non-myristoylated Gα_{i1} are very similar, however, when myristoylation has taken place on the N-terminus of Gα_{i1}, the conformation of the subunit changes drastically, leading

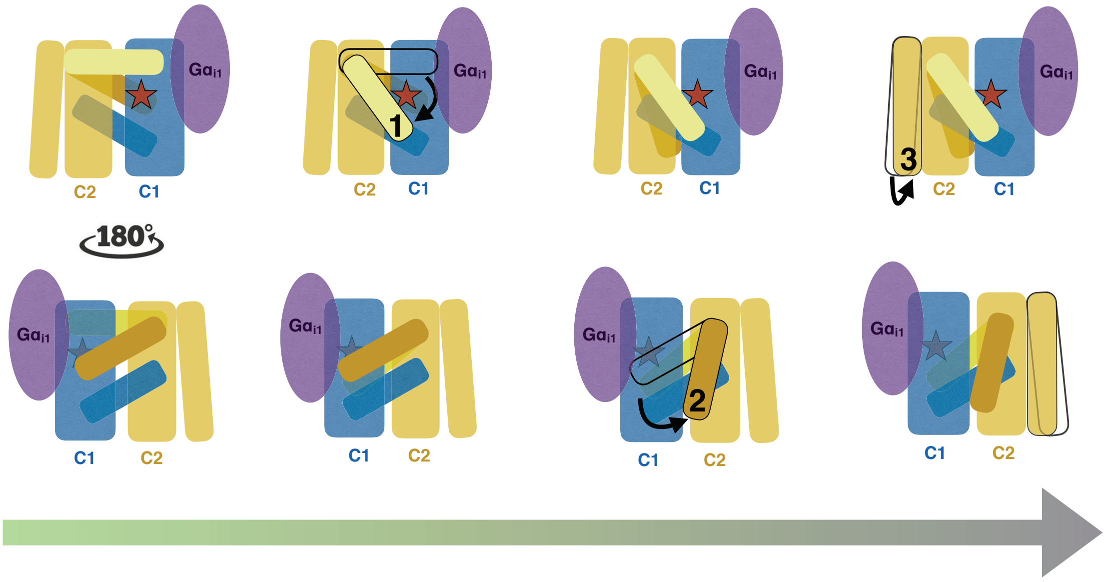


Figure 5.7: Graphical representation of proposed AC5 inhibition mechanism by $G\alpha_{i1}^{myr}$ with the upper row showing the cytosolic side of AC5 and the lower row depicting AC5 from the membrane side. The change that takes place in step (1) compared to the initial close-to-active AC5 conformation, is the relocation of the C2 $\beta 7$-$\beta 8$ loop away from its active position. This alteration takes place near AC5's active site (red star), which is also affected by this event. Conformational change (2) involves the loss of interaction between C1's $\alpha 2$ and the C2 $\beta 4$-$\beta 5$ loop, weakening the active site. The final rearrangement (3) includes the closer packing of C2's $G\alpha_s$ interaction site, which appears to result in a less favourable C2 conformation for the interaction with $G\alpha_s$.

to a structure that differentiates itself from the active Gα_s and that enables the protein to function in an inhibitory fashion as is shown via the presented classical MD simulations. Hence, in line with experimental studies, myristoylation appears to be crucial for G$_i$'s function and demonstrates how important even relatively small changes to a protein structure can be for its function.

6 Conclusions and Outlook

The first step that takes place in the GPCR signal transduction pathway is the transfer of an extracellular signal to the intracellular side. In this thesis this signal transduction across the membrane was explored as well as further transmission of the signal to other proteins in the cytosol, Gα_i and AC. The study of the GPCR rhodopsin at body temperature showed that the protein environment of the active site plays a significant role in biasing the chromophore into its active configuration after photon exposure. Besides the shape of the active site, also the rearrangements in the hydrogen bond network and the presence of water molecules in the active site appear to be crucial for signal propagation from the chromophore to the protein environment. Because water molecules are not covalently bound to the protein, they increase the mobility of the active site and can therefore enhance the adaptability as well as assist structural rearrangements required during photo activation. Rhodopsin's activation mechanism demonstrates the importance of the conformation of the active site, but also reflects the sensitivity of the protein to temperature as different photo activation pathways have been experimentally observed at low and room temperature. Protein crystallisation procedures can influence rhodopsin's conformation as well and can significantly alter the state of the GPCR or the presence of water molecules in the protein.

The presented simulations emphasise that when a protein is sensitive to temperature fluctuations this should be taken into account when investigating reaction pathways using structures obtained via X-ray crystallisation. However, the neutral Schiff base rhodopsin state, obtained via QM/MM MD at body temperature, does not reach a fully adapted rhodopsin conformation yet at the end of the QM/MM MD trajectory. This is because QM/MM MD simulations of the system can only reach the picosecond time scale while the relaxation of the active site and the rest of the protein will take place on the nanosecond time scale. Therefore, QM/MM

MD simulations are not suitable to study the effect of the active site on the other regions of the protein. The development of a neutral *all*-trans-retinal force-field parameter set could facilitate the switch from QM/MM MD to classical MD, which would allow the protein system to relax on the microsecond time scale. Such MD simulations of the neutral *all*-trans retinal in rhodopsin would provide the opportunity to study the further relaxation of the active site, the propagation of the change in retinal structure to other regions of the protein and the role of water molecules during rhodopsin activation to the meta II form. Another approach that could be used is multiple time-step QM/MM:MM MD, currently implemented by Rothlisberger's laboratory, in which fast components of the forces are calculated via force-field based MD and the difference in force between the QM forces and the MM forces is integrated with a long time step. In this way first-principle MD simulations can be performed in which the time step can be increased to a similar level as MM MD. The application of this approach to the deprotonated Schiff base conformation could provide a detailed description of the further relaxation of the active site around the Schiff base, which might not be possible with the direct application of classical MD after QM/MM MD. In order to go to very long time scales, enhanced sampling techniques could be employed to attempt to sample conformations of the fully activated rhodopsin system in which the ionic lock between Arg135 and Glu247 is broken.

The second step of the GPCR signal transduction pathway that was covered in this thesis is the activation of the inhibitory Gα subunit. Although N-terminal myristoylation of Gα_i is commonly excluded from the crystallisation process, the classical MD simulations suggest that the addition of the lipid to the protein structure has an impact over a large region of the protein surface. Besides the formation of a hydrophobic pocket in the *Ras* domain, also other regions of the subunit are altered, which play an essential role in the protein's function. Hence, the performed simulations suggest that nature uses N-terminal myristoylation to induce specificity and to regulate the function of closely related G proteins. The presence or absence of a small post-translational modification could be viewed as a minimal effort to tune the same building blocks for diverse (even inverse) functions in the cell.

The study of G-protein AC regulation, the third study included in this thesis, was part of a sub group in the Human Brain Project in which several European laboratories, specialised in systems biology, Brownian dynamics, classical MD and QM/MM MD, collaborated to obtain a better understanding of the rewarding-memory cascade or reinforcement learning. A pathway in the brain that, due to previous encounters, can associate an external stimulus to a reward or a reward-producing response [Nair et al., 2015]. The systems biology model of the pathway includes a biochemical network that starts with GPCR stimulation and ends with protein kinase A activation [Nair et al., 2015]. The presence of a Gα_{i1}^{myr}:AC5:Gα_s complex – of which its formation is unclear as it has never been reported in experimental studies [Dessauer et al., 1998] – or solely Gα_{i1}^{myr}:AC5 and Gα_s:AC5 complexes has a significant effect on AC regulation in the modelled pathway. In order to obtain a better understanding of Gα_{i1}^{myr}'s inhibitory effect on AC, a classical MD study of the Gα_{i1}^{myr}:AC5 complex in comparison with the free AC5 system was performed. According to the MD simulations, the interface between AC5's C1 and C2 domain appears to be the main region that is affected by the presence of the Gα subunits (Gα_{i1}^{myr} as well as Gα_s). Gα_s can stabilise C2's β7-β8 loop via its interactions with the C2 domain, which assists the formation of the ATP binding site. On the other hand, Gα_{i1}^{myr} stabilises C2's β7-β8 loop in a location that seems to interfere with ATP binding. In addition, Gα_{i1}^{myr}'s complexation to AC5 leads to a loop relocation in the C2 domain, increasing the perturbation of the active site. The presence of Gα_{i1}^{myr} also appears to disfavour possible Gα_s binding to the C2 domain, suggesting a double mechanism of inhibition via the ATP binding site and the Gα_s binding site. Moreover, the simulations of Gα_{i1}^{myr}:AC5 and free AC5 indicate that despite the high structural similarity of the Gα_s and non-myristoylated Gα_{i1} X-ray structures, Gα_{i1}^{myr} and Gα_s appear to act on two different sites on AC's catalytic domains.

The mechanistic insights obtained via the investigation of soluble Gα_{i1}^{myr} and AC's interaction with Gα subunits could provide a new perspective on Gα subunit regulation. Before this study was conducted, the stimulatory and inhibitory Gα subunits appeared to be structurally very similar according to the X-ray structures of active Gα_s and non-myristoylated Gα_i. However, due to the adjustments in Gα_i's conformation after N-myristoylation, the subunits Gα_i^{myr} and Gα_s show distinct structural differences, that seem to be essential for differentiating their action. In combination with experimental studies, the simulation results also suggest the possibility of the presence of different Gα_i conformations in the active (GTP bound) state that are either soluble when the myristoyl moiety is in the protein's hydrophobic pocket or can interact with the membrane, which leads to a relocation of the myristoyl group away from the hydrophobic pocket and towards the membrane. Hence, the inhibition via Gα_i subunits could be more complicated than previously assumed as at least two conformations of the Gα subunit can be present on the intracellular side.

The better understanding of the possible interactions between Gα_{i1}^{myr} and AC5 and the inhibitory effect of Gα_{i1}^{myr} on AC5 obtained in this thesis, helped to identify new key regions of the enzyme that are essential for activation or inhibition and suggested a new view on G-protein induced conformational changes of AC5. This new perspective on Gα subunit regulation impacts besides our understanding of signal transduction also drug development that targets the GPCR signal transduction pathway as the structural difference between Gα_i and Gα_s could open up new avenues for enhancing/reducing AC's interaction with Gα_i or Gα_s.

Although these first simulations of Gα_{i1}^{myr}, AC5 and Gα_{i1}^{myr}:AC5 suggest many interesting new perspectives, these findings should also be verified via experimental studies. One approach to test the exact location of the myristoyl moiety could be to mutate the hydrophobic binding site in such a way that the myristoyl moiety is not able to fit in the pocket anymore, for instance, via the mutation of Met53 and Phe196 into bulky residues. Because it is known that non-myristoylated Gα_{i1} is unable to bind to AC's C1, the interaction between C1 and Gα_{i1}^{myr} with the mutated hydrophobic pocket could be tested to investigate if the binding of the myristoyl moiety in the hydrophobic pocket directly affects the interaction with C1. Another experiment that could be considered to study the Gα_{i1}^{myr}/AC5 interface is a FRET experiment between Gα_{i1}^{myr} and AC5 for which a donor and acceptor group are attached to the alpha helical domain of Gα_{i1}^{myr} and the C1 domain of AC5.

Besides experimental studies also several additional simulations can be conducted to improve the understanding of Gα regulation. In this thesis the free AC5 system and the Gα_{i1}^{myr}:AC5 complex were investigated starting from an AC template of substrate-free Gα_s:AC. However, many alternative states of AC still remain to be tested as, for example, the stability of the Gα_{i1}^{myr}:AC5:Gα_s complex. Currently, the results of Gα_{i1}^{myr}:AC5 and free AC5 suggest that Gα_{i1}^{myr} can act as a preventer of ATP and Gα_s association. If this is the case, the AC conformation that is initially interacting with an analog of ATP and Gα_s could be affected less by the presence of Gα_{i1}^{myr} because ATP induces a conformation of C1 that appears to be less favourable for Gα_{i1}^{myr} association. Therefore, complexes of Gα_{i1}^{myr}:AC5 and Gα_{i1}^{myr}:AC5:Gα_s with PDB code 1CJK (Gα_s:AC bound to an ATP analog) as AC5 template and Gα_{i1}^{myr}:AC5 and Gα_{i1}^{myr}:AC5:Gα_s with PDB code 3C15 (Gα_s:AC bound to pyrophosphate) as AC5 template have also to be included in followup studies to further validate the proposed inhibition mechanism. Via the simulation of these different systems it could be investigated how both Gα subunits influence each other's interaction with AC5 and thus modify AC5's function. In addition, these simulations could show the effect of different AC5 conformations on the stability of the Gα interactions with the enzyme. Overall, the results could reveal the influence of Gα subunits on AC5, depending on different states of AC5 activation, the possibility of Gα_{i1}^{myr}:AC5:Gα_s formation as well as lead to a better understanding of which AC5 conformations can be inhibited by Gα_{i1}^{myr}.

A How Rhodopsin Tunes the Equilibrium between Protonated and Deprotonated Forms of the Retinal Chromophore

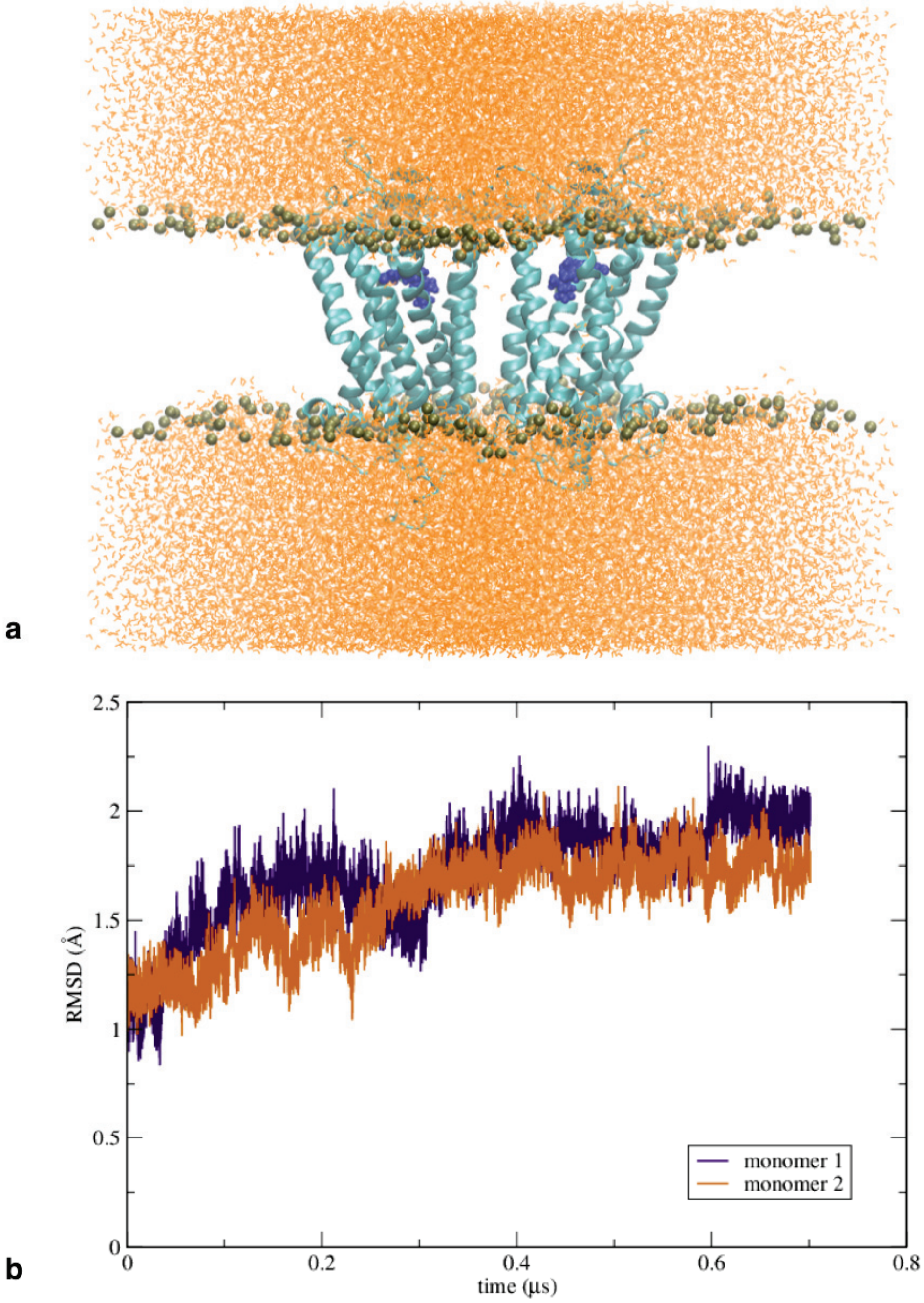


Figure A.1: Rhodopsin dimer system setup. (a) Rhodopsin dimer system in which water is depicted in orange, the phosphor atoms in the lipids are shown in tan and the protein dimer is represented in cyan with the retinal moiety in dark blue. (b) Root mean square deviation of the backbone of both rhodopsin monomers including residue 2 to 322. The structure used as a reference structure for the RMSD was the dark state dimer structure obtained from combining both 1U19 and 1GZM structures. Monomer 2 was used to perform the *cis*-*trans* isomerisation on.

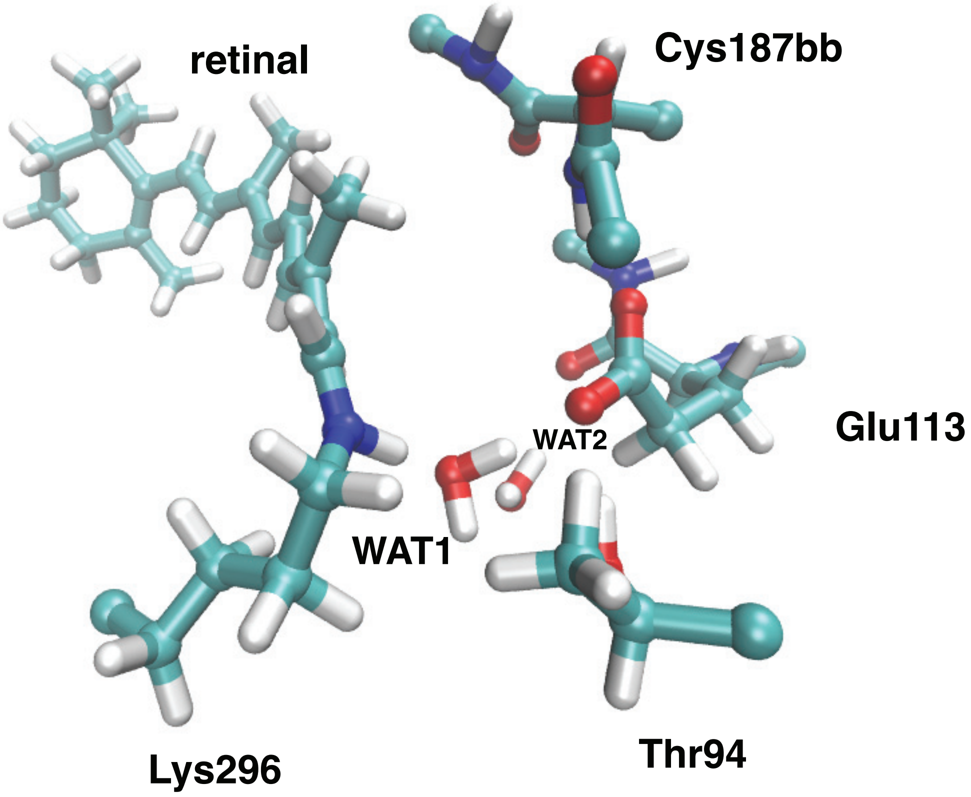


Figure A.2: QM region that is used in the thermodynamic integration simulations. To the final carbons, before the cut with the protein environment, monovalent pseudo potentials were applied.

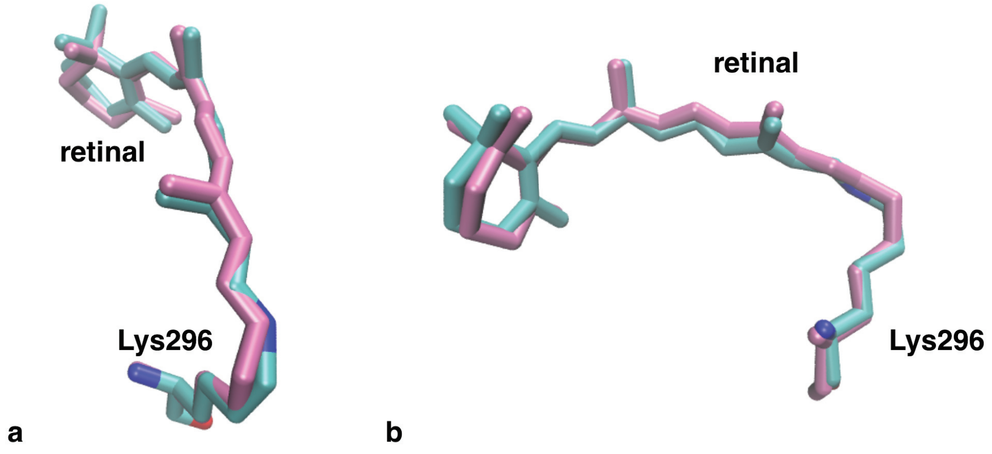


Figure A.3: Alignment of the retinal and Lys296 moieties of the batho conformation from the QM/MM MD simulation in pink and the batho X-ray structure in cyan (carbon), blue (nitrogen) and red (oxygen) (PDB code 2G87). The structure of the retinal and Lyr296 are represented without hydrogens for the QM/MM MD simulations. (a) and (b) show two different perspectives from the PSB and the β-ionone ring. The alignment of the retinal moiety and the protonated Schiff base of both batho structures (from the QM/MM simulations and the X-ray structure) results in a RMSD of 0.36 Å. The RMSD between QM/MM batho and dark state structures is 0.57 Å when the retinal and the protonated Schiff base are aligned.

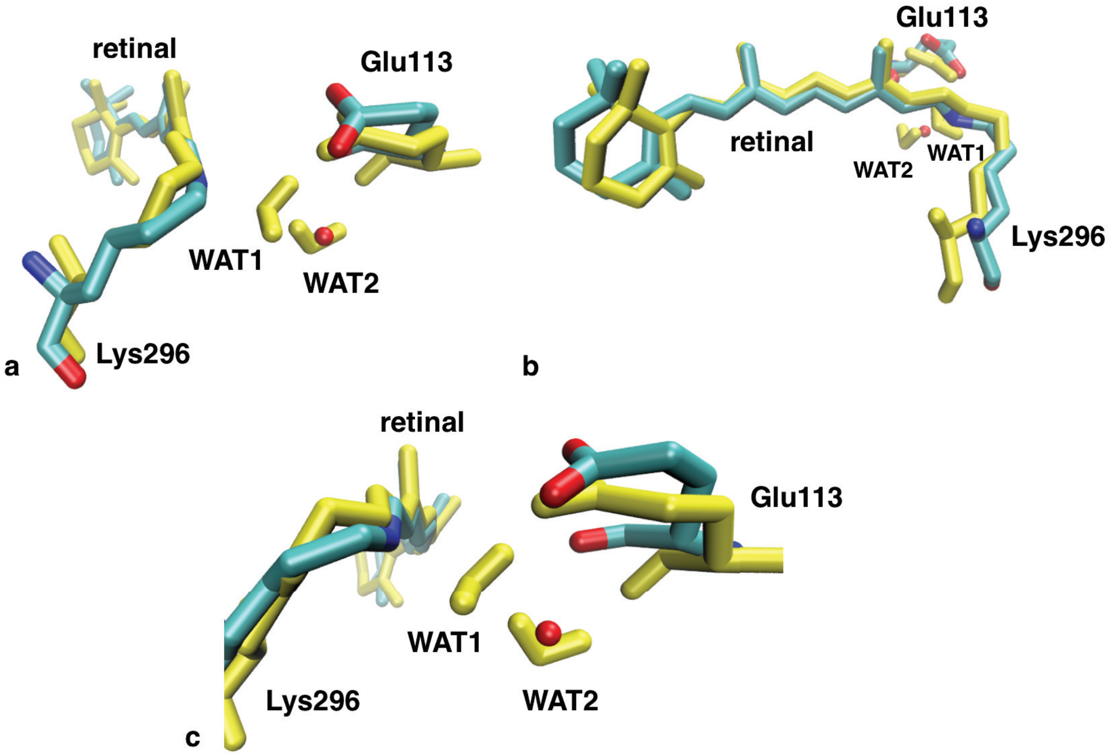


Figure A.4: Alignment of the retinal and Lys296 moieties of the lumi conformation from the QM/MM MD simulation in yellow and the lumi X-ray structure in cyan (carbon), blue (nitrogen) and red (oxygen) (PDB code 2HPY). The structure of the retinal, Lys296 and Glu113 are represented without hydrogens for the QM/MM MD simulations, but for the water molecules the hydrogens are depicted. (a) and (b) show two different perspectives from the PSB and the β-ionone ring. The alignment of the retinal moiety and the protonated Schiff base of both lumi structures (from the QM/MM simulations and the X-ray structure) results in a RMSD of 0.51 Å. The RMSD between QM/MM lumi and dark state structures is 0.70 Å when the retinal and the protonated Schiff base are aligned. (c) provides a zoom of the indirect hydrogen bond between Glu113 and PSB in the X-ray and the QM/MM structure.

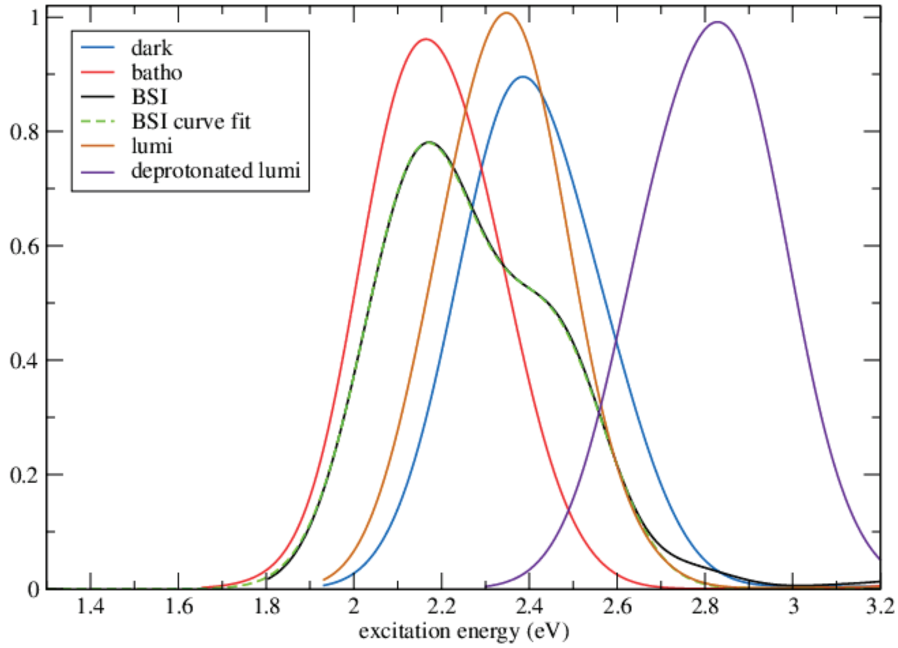
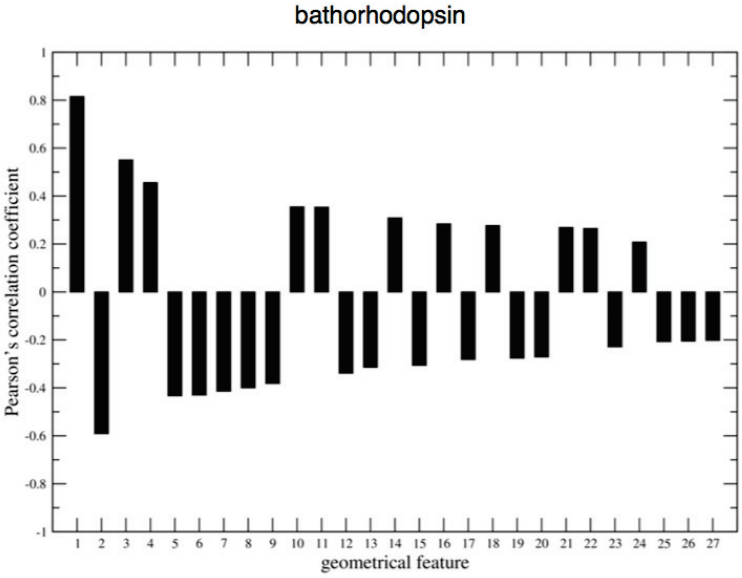


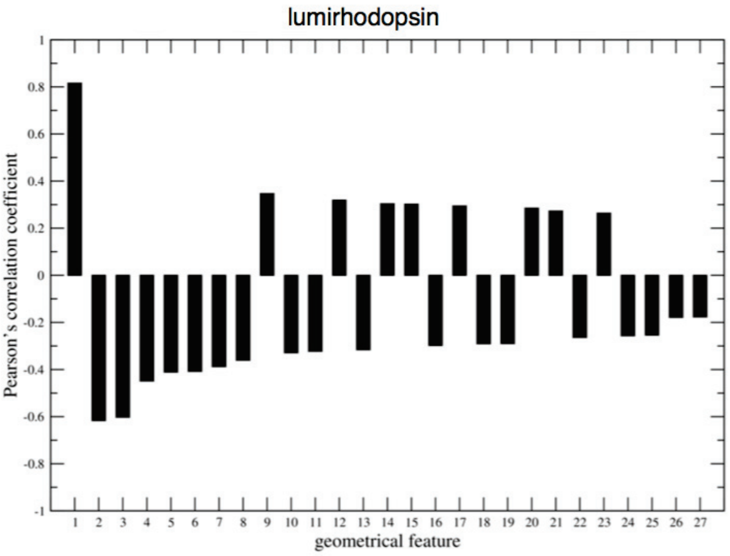
Figure A.5: Excitation energies generated via ZINDO/S calculations of the dark state and several other intermediate states of rhodopsin present after *cis*-*trans* isomerisation.

Figure A.6: Correlation profiles of bathorhodopsin, lumirhodopsin and pre-metarhodopsin I.



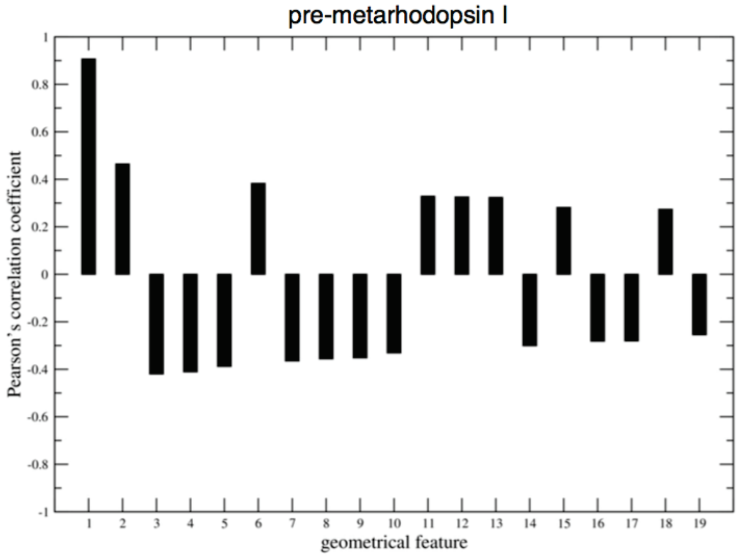
1. BLA
2. d(C9-C10)
3. α(C8-CZPhe208-C9)
4. α(C8-CE2Phe91-C9)
5. α(C5-CE2Phe293-C6)
6. γ(C18-CG2Thr118-CBTyr268-CGMet288)
7. γ(CG1Ile189-C10-C12-C14)
8. d(C11-C12)
9. γ(CD1Leu125-C8-C9-C13)
10. γ(C1-CE2Phe261-CD1Phe293-CGPhe293)
11. γ(CE1Phe91-CBGlu181-OE2Glu181-CE1Phe208)
12. γ(CGPhe91-CD2Phe91-CBTyr178-CZ3Trp265)
13. γ(CBPhe261-CGTrp265-CGMet288-CE2Tyr268)
14. γ(C8-CGPhe91-CEMet288-CGMet288)
15. γ(CGTyr178-CZ3Trp265-CE3Trp265-CZPhe91)
16. γ(C6-C8-C17-C9)
17. γ(CBThr94-CDGlu113-CE1Tyr178-CE1Tyr191)
18. γ(OHTyr191-CZ2Trp265-CZTyr268-CEMet288)
19. γ(C17-CD1Phe208-CE3Trp265-CGTrp265)
20. γ(CGGlu181-C13-C14-C3)
21. γ(OG1Thr94-CG2Thr118-OGSER186-OE1Glu181)
22. α(CE2Phe212-CD2Phe91-CD2Phe212)
23. γ(CD1Tyr268-C8-C11-C10)
24. α(CBPhe212-CE2Phe91-CGPhe212)
25. γ(C20-CGMet44-CD1Phe208-CGTrp265)
26. d(CD1Tyr178-CD2Tyr178)
27. γ(CGPhe91-CDGlu181-CZ3Trp265-CE2Phe212)

d = distance
α = angle
γ = dihedral



1. BLA
2. d(C11-C12)
3. d(C9-C10)
4. α(C11-CBPhe261-C12)
5. γ(CE2Tyr191-C13-C19-C12)
6. γ(C17-SDMet44-CBPhe208-OE1Glu181)
7. α(C9-NE1Trp265-C10)
8. γ(C14-CGGlu113-CD1Tyr268-SDMet288)
9. α(C8-CG1Ile189-C9)
10. γ(CDGlu113-CE1Phe208-CD2Tyr268-CZPhe208)
11. γ(CGGlu181-CBSER186-CE2Tyr268-CGPhe261)
12. γ(C7-SDMet44-CE1Tyr191-OE2Glu181)
13. γ(CD1Phe91-CE1Phe91-CH2Trp265-CD2Phe212)
14. γ(CGPhe212-CBAla269-CE2Phe293-CD1Phe293)
15. γ(CD2Phe91-SGCYS167-CZTyr191-CE1Tyr191)
16. γ(CZPhe208-CBAla272-OGSER298-CGPhe293)
17. γ(CD1Ile189-CBPhe208-CE2Phe293-CD2Phe293)
18. γ(CGLeu125-CGPhe208-CE3Trp265-CZPhe208)
19. γ(OHTyr268-CE2Tyr268-CBAla269-CDGlu113)
20. γ(CBMet44-OE1Glu113-OE2Glu113-CD2Tyr178)
21. γ(C4-OE2Glu181-CZTyr268-SDMet288)
22. γ(C10-CEMet44-OE2Glu181-CBMet44)
23. γ(CD2Phe91-OG1Thr94-CBGlu113-CGGlu113)
24. γ(C1-CE1Tyr191-CD2Trp265-CE2Tyr268)
25. γ(C11-C12-C20-C13)
26. γ(CG2Ile189-CGTyr268-CZTyr268-CE1Tyr268)
27. γ(C15-CGPhe261-CD2Trp265-CE3Trp265)

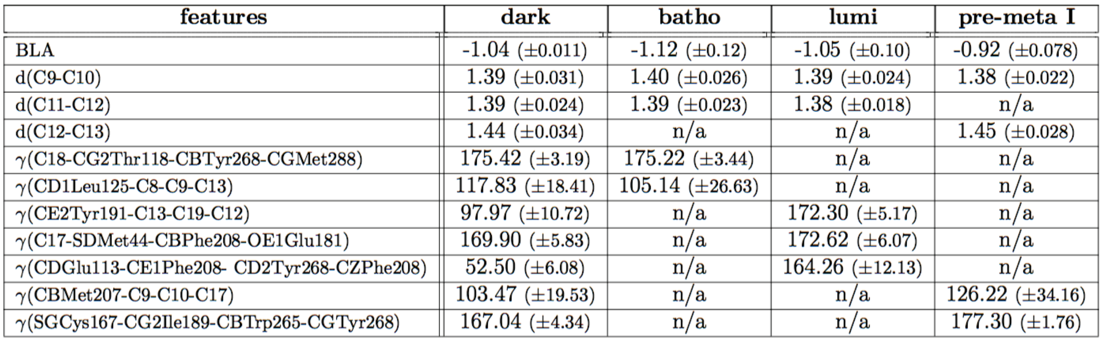
d = distance
α = angle
γ = dihedral



1. BLA
2. d(C12-C13)
3. γ(CBMet207-C9-C10-C17)
4. d(C9-C10)
5. γ(C3-CEMet44-CD1Phe261-CGPhe261)
6. γ(C4-CD1Tyr191-CZPhe261-CBSer298)
7. γ(C11-CBMet44-CGMet207-CBPhe293)
8. γ(SGCys167-CG2Ile189-CBTrp265-CGTyr268)
9. γ(C14-CD1Phe91-CGTyr191-CBTyr191)
10. γ(C13-OE2Glu113-CE3Trp265-CBSer298)
11. γ(CD1Tyr191-C9-C20-C10)
12. α(C14-C13-C15)
13. γ(CE1Tyr268-CZTyr268-CBPhe293-CD1Phe91)
14. γ(SDMet86-CG2Thr118-CGPhe293-CD1Phe293)
15. γ(OE1Glu181-CZTyr268-CBAla272-C14)
16. γ(C2-CZPhe91-CZ2Trp265-CE2Phe91)
17. γ(CGPhe208-CGTyr268-CBAla292-CE2Phe212)
18. γ(CD2Phe261-CZTyr268-CE2Tyr268-CD2Tyr191)
19. γ(CBTyr191-CGTrp265-CD2Trp265-CH2Trp265)

d = distance
α = angle
γ = dihedral

Table A.1: Mean and standard deviations of features that are included in the causality profiles of batho, lumi and pre-meta I. The BLA and distances are depicted in Ångstrom, the dihedral angles are shown in degrees. The values of the features are depicted when the feature is part of the intermediate's causality scheme, otherwise the mean and standard deviations are not applicable. The mean and standard deviations for all features are shown for the reference state, the dark conformation.



features	dark	batho	lumi	pre-meta I
BLA	-1.04 (±0.011)	-1.12 (±0.12)	-1.05 (±0.10)	-0.92 (±0.078)
d(C9-C10)	1.39 (±0.031)	1.40 (±0.026)	1.39 (±0.024)	1.38 (±0.022)
d(C11-C12)	1.39 (±0.024)	1.39 (±0.023)	1.38 (±0.018)	n/a
d(C12-C13)	1.44 (±0.034)	n/a	n/a	1.45 (±0.028)
γ(C18-CG2Thr118-CBTyr268-CGMet288)	175.42 (±3.19)	175.22 (±3.44)	n/a	n/a
γ(CD1Leu125-C8-C9-C13)	117.83 (±18.41)	105.14 (±26.63)	n/a	n/a
γ(CE2Tyr191-C13-C19-C12)	97.97 (±10.72)	n/a	172.30 (±5.17)	n/a
γ(C17-SDMet44-CBPhe208-OE1Glu181)	169.90 (±5.83)	n/a	172.62 (±6.07)	n/a
γ(CDGlu113-CE1Phe208- CD2Tyr268-CZPhe208)	52.50 (±6.08)	n/a	164.26 (±12.13)	n/a
γ(CBMet207-C9-C10-C17)	103.47 (±19.53)	n/a	n/a	126.22 (±34.16)
γ(SGCys167-CG2Ile189-CBTrp265-CGTyr268)	167.04 (±4.34)	n/a	n/a	177.30 (±1.76)

Figure A.7: Graphs of distances of bonds and hydrogen bonds in the unconstrained deprotonated lumi configuration, after performing TI on the lumi system. The distances in the graphs are visualised via coloured highlights in a schematic representation of the active site in the lumi state. The coloured regions corresponds to the colour code of the distances in the graphs.

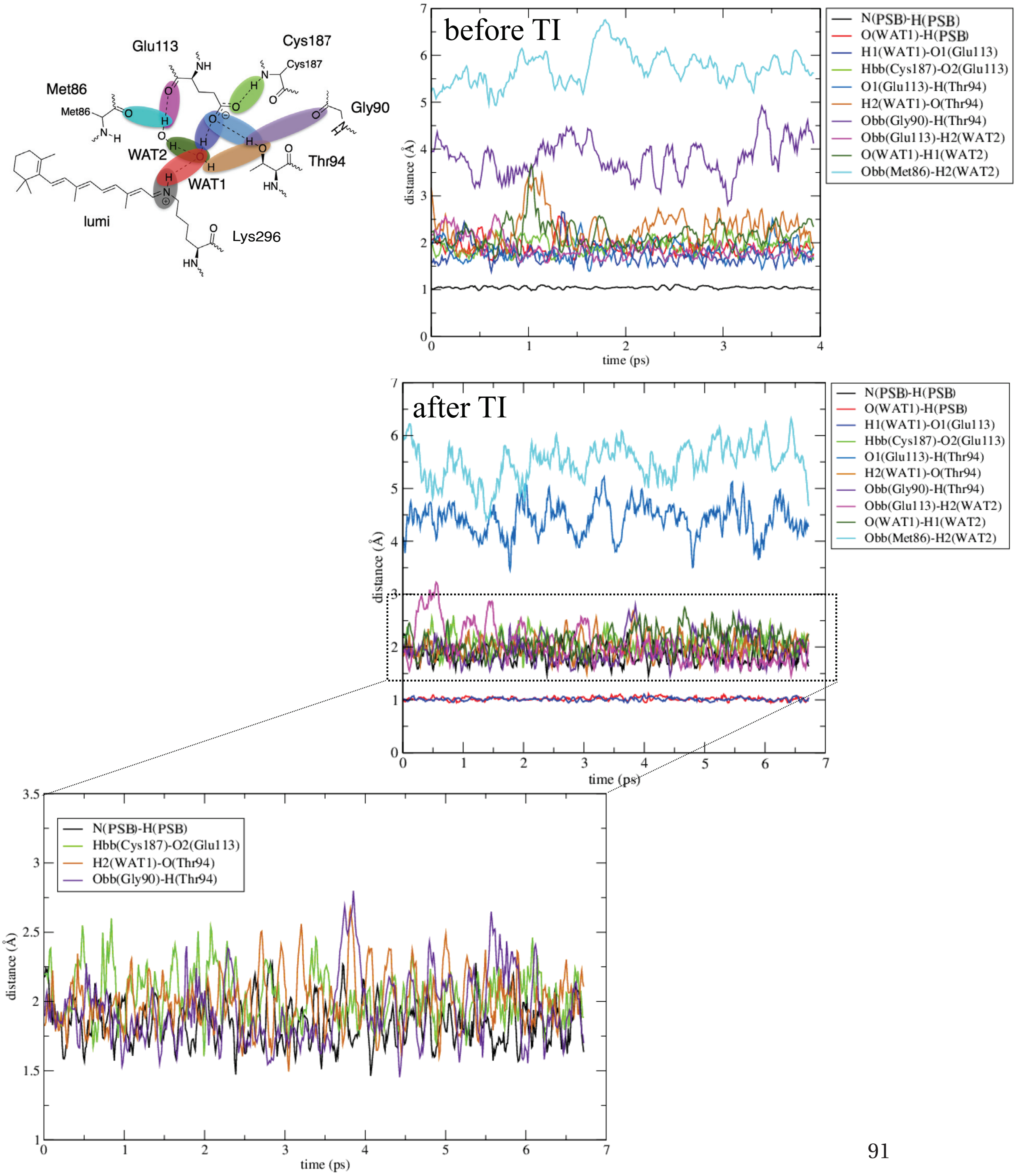
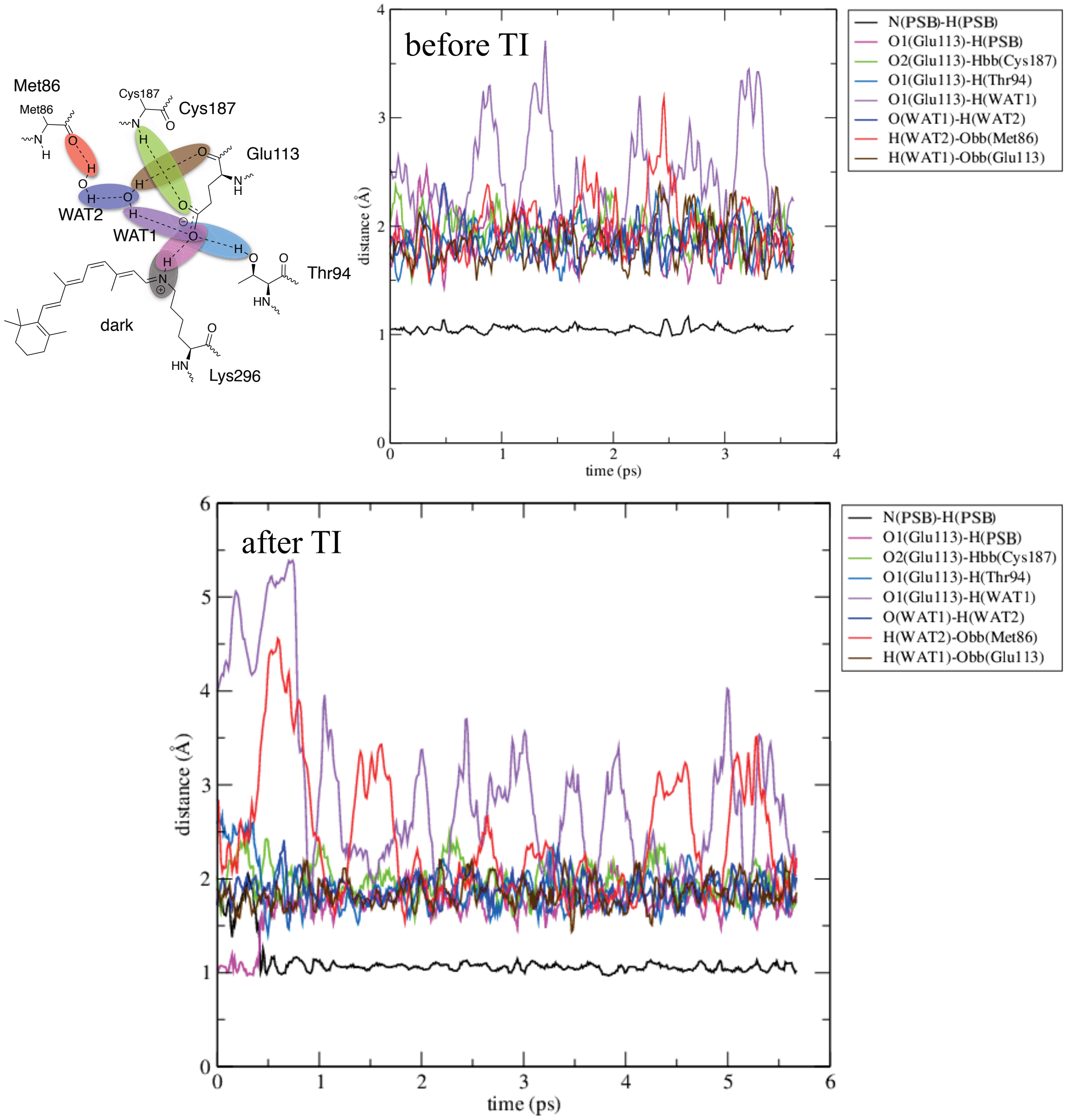


Figure A.8: Graphs of distances of bonds and hydrogen bonds in the unconstrained protonated dark state before TI and the deprotonated dark configuration, after performing TI on the dark system. The distances in the graphs are visualised via coloured highlights in a schematic representation of the active site in the dark state. The coloured regions correspond to the colour code of the distances in the graphs.



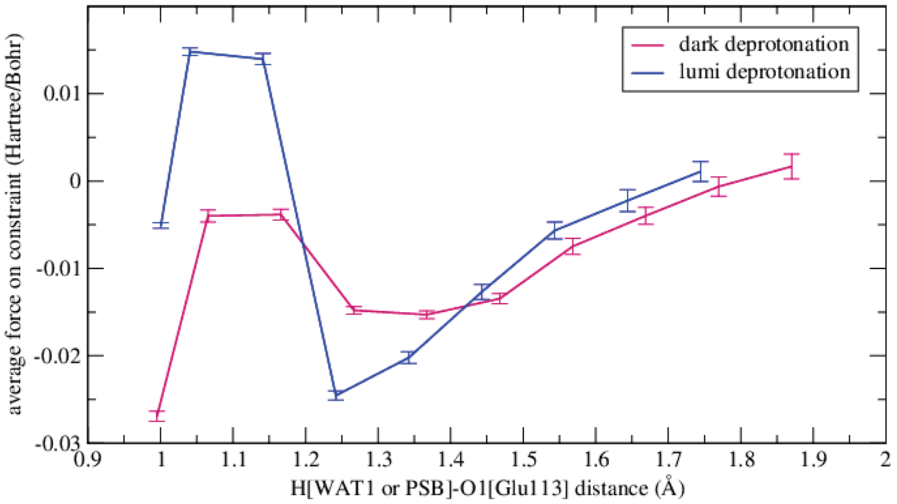


Figure A.9: Average force on constraint in Hartree/Bohr for every constraint window. For every window the standard deviation is shown as well.

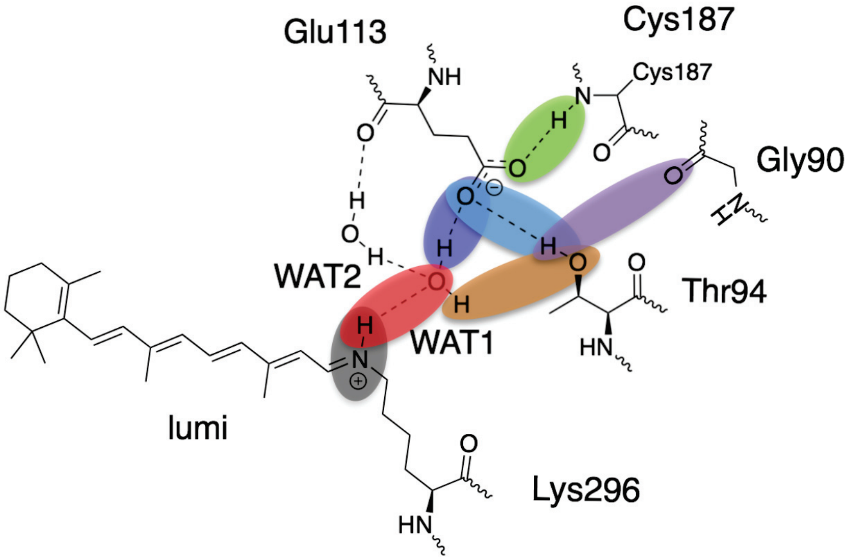
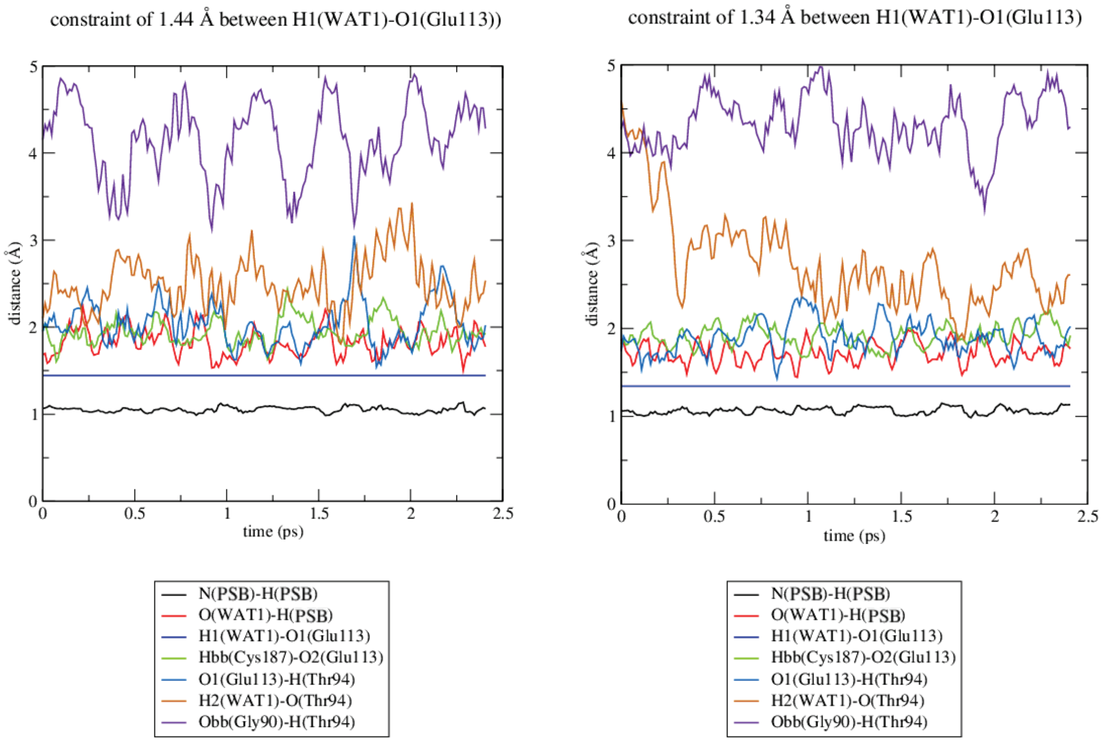
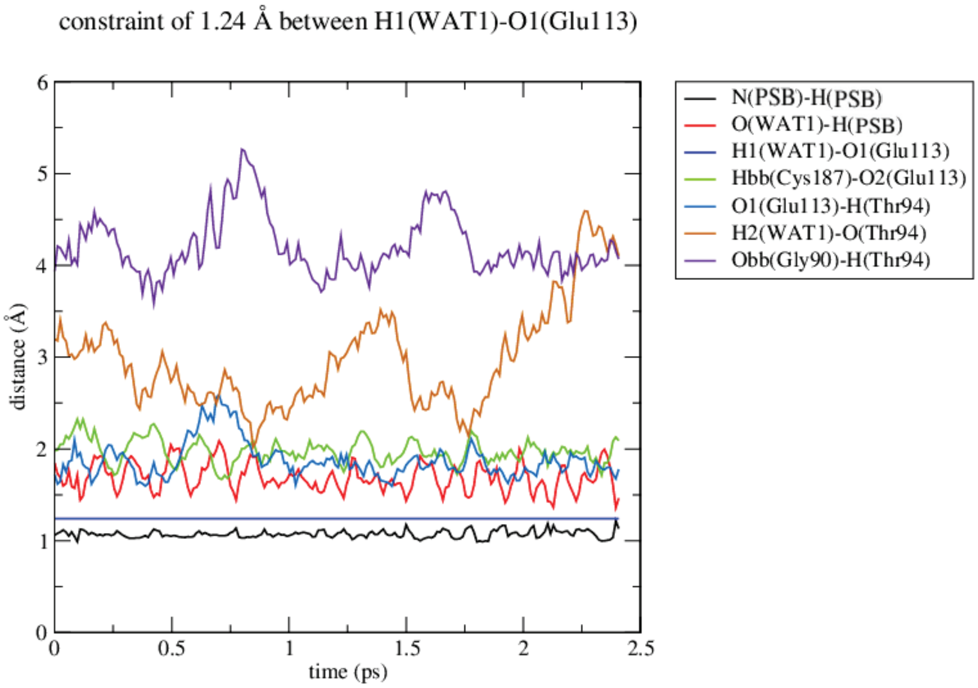
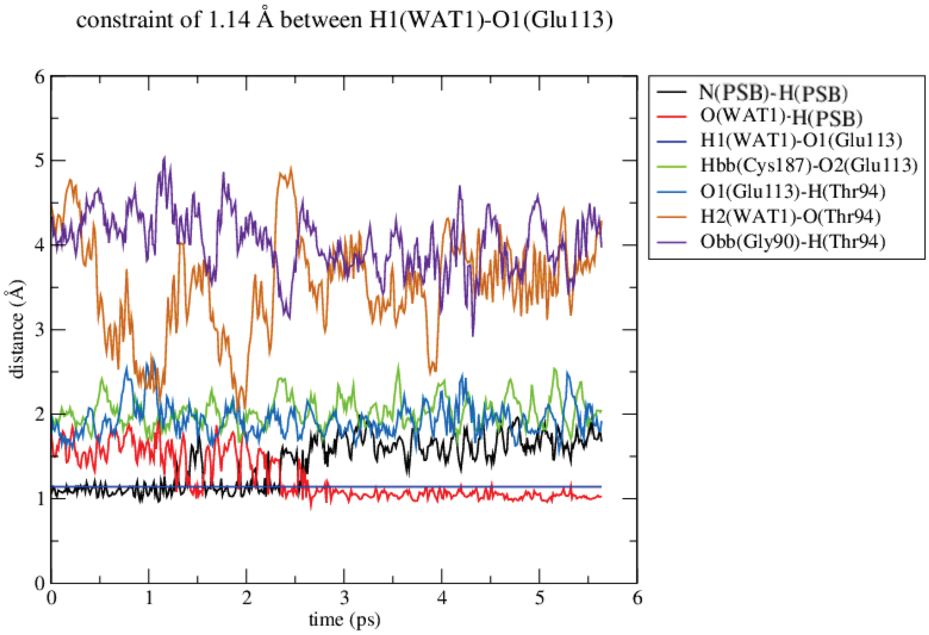


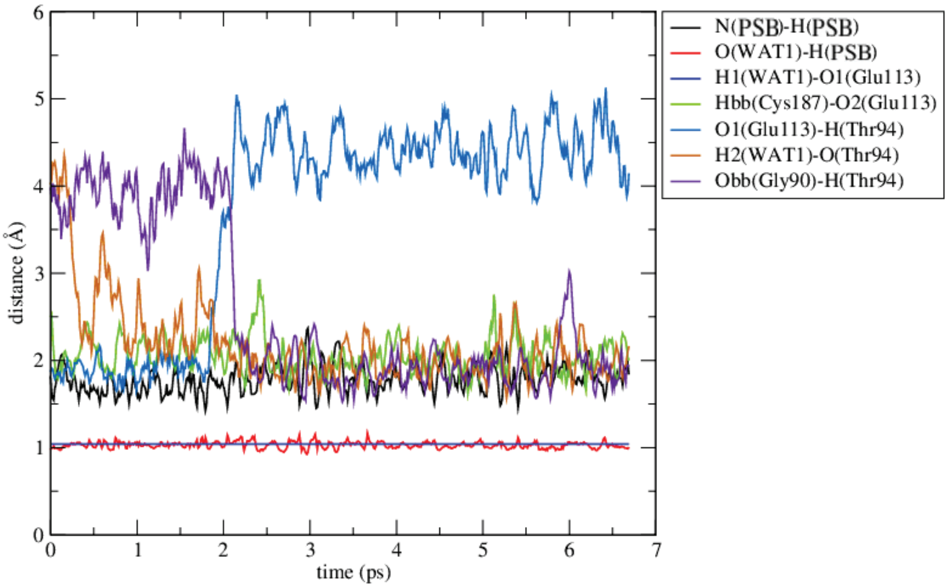
Figure A.10: Colour assignment of the covalent bond and hydrogen bonds that were used to monitor the change in environment of PBS upon deprotonation in the lumi state (Fig. A.11).

Figure A.11: Graphs of distances of bonds and hydrogen bonds at different distances (1.44 Å, 1.34 Å, 1.24 Å, 1.14 Å, 1.04 Å, 1.00 Å) of the constraint between H1(WAT1) and O1(Glu113) during the TI calculation in the lumi state.

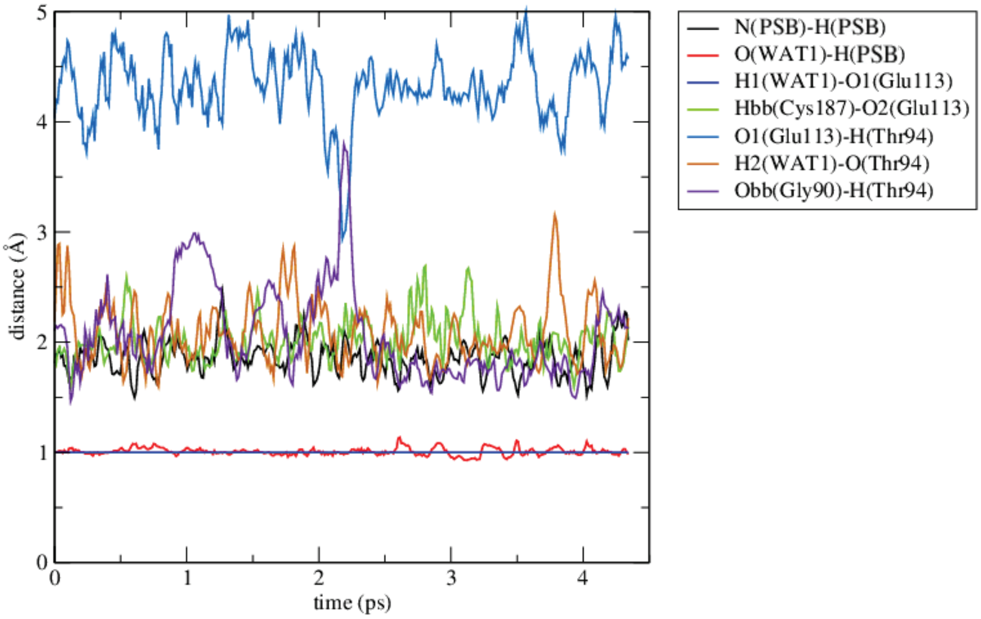



constraint of 1.24 Å between H1(WAT1)-O1(Glu113)
distance (Å)
0
1
2
3
4
5
6
time (ps)
0
0.5
1
1.5
2
2.5
N(PSB)-H(PSB)
O(WAT1)-H(PSB)
H1(WAT1)-O1(Glu113)
Hbb(Cys187)-O2(Glu113)
O1(Glu113)-H(Thr94)
H2(WAT1)-O(Thr94)
Obb(Gly90)-H(Thr94)


constraint of 1.14 Å between H1(WAT1)-O1(Glu113)
distance (Å)
0
1
2
3
4
5
6
time (ps)
0
1
2
3
4
5
6
N(PSB)-H(PSB)
O(WAT1)-H(PSB)
H1(WAT1)-O1(Glu113)
Hbb(Cys187)-O2(Glu113)
O1(Glu113)-H(Thr94)
H2(WAT1)-O(Thr94)
Obb(Gly90)-H(Thr94)



constraint of 1.04 Å between H1(WAT1)-O1(Glu113)



constraint of 1.00 Å between H1(WAT1)-O1(Glu113)

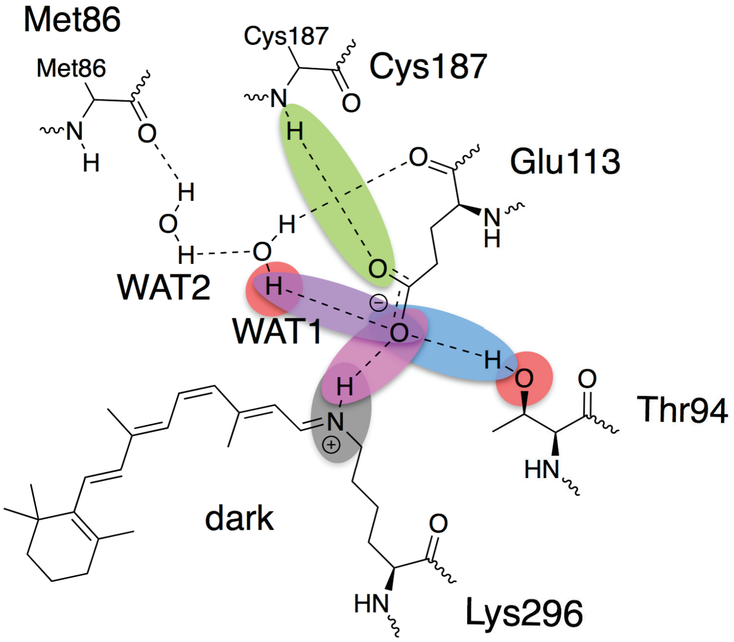
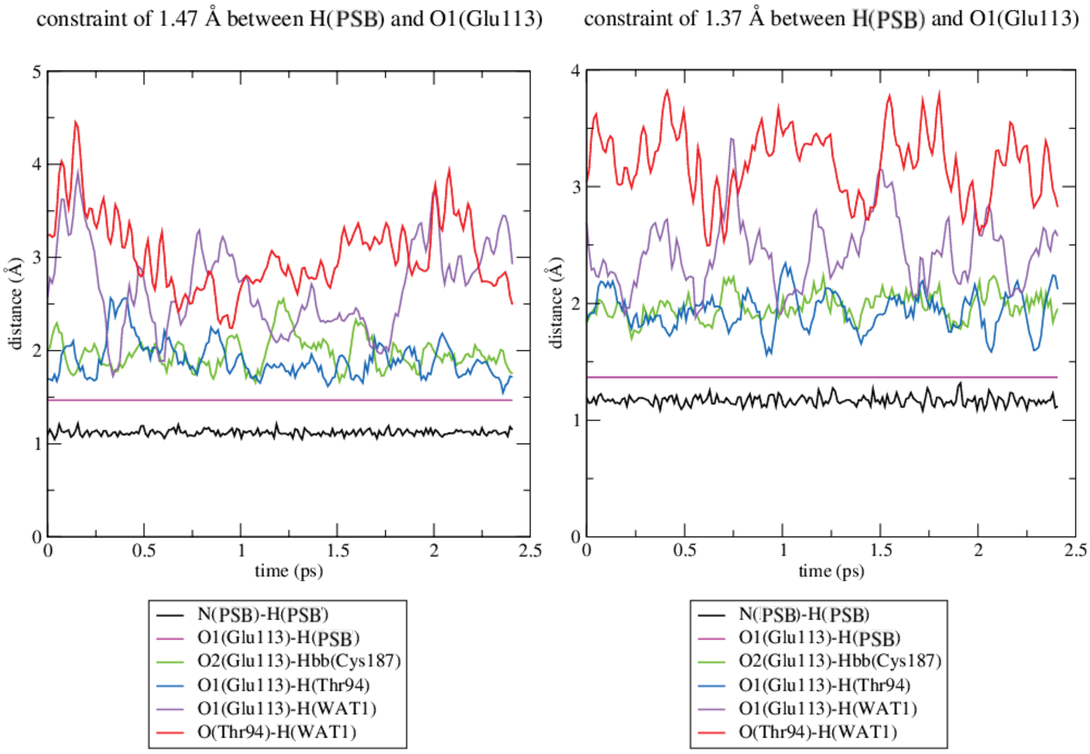


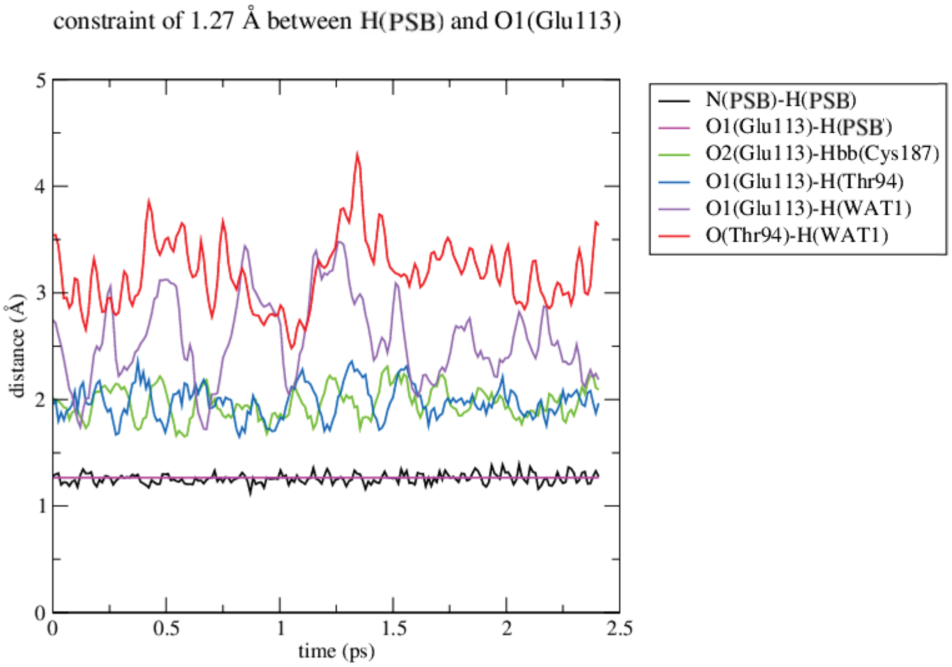
Figure A.12: Colour assignment of the covalent bond and hydrogen bonds that were used to monitor the change in environment of PBS upon deprotonation in the dark state (Fig. A.13).

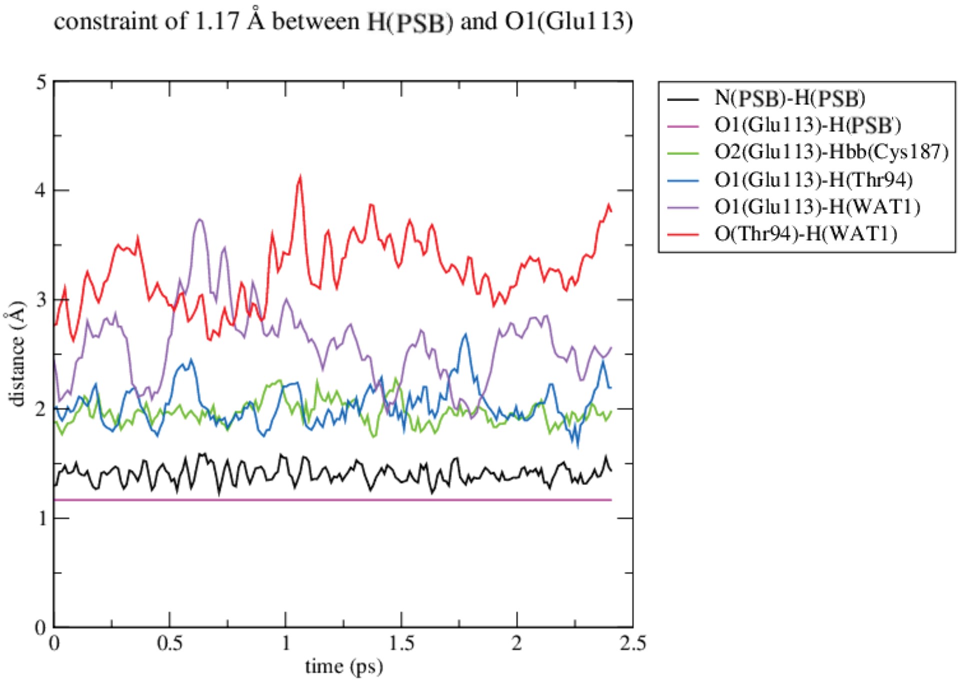
Figure A.13: Graphs of distances of bonds and hydrogen bonds at different distances (1.47 Å, 1.37 Å, 1.27 Å, 1.17 Å, 1.07 Å, 1.00 Å) of the constraint between H(PSB) and O1(Glu113) during the TI calculation in the dark state.



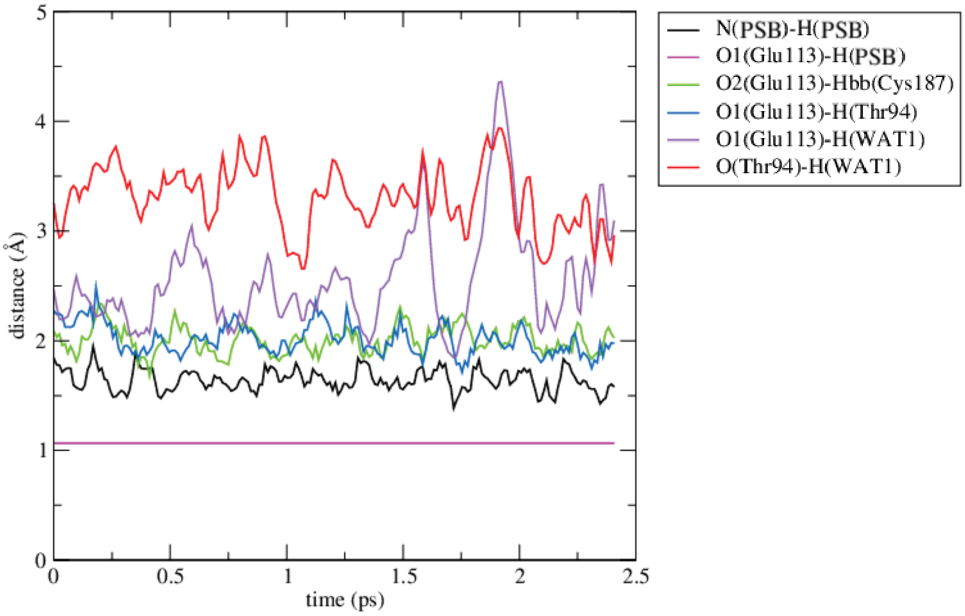
constraint of 1.47 Å between H(PSB) and O1(Glu113)

constraint of 1.37 Å between H(PSB) and O1(Glu113)


constraint of 1.27 Å between H(PSB) and O1(Glu113)
distance (Å)
time (ps)
0
1
2
3
4
5
0
0.5
1
1.5
2
2.5
N(PSB)-H(PSB)
O1(Glu113)-H(PSB)
O2(Glu113)-Hbb(Cys187)
O1(Glu113)-H(Thr94)
O1(Glu113)-H(WAT1)
O(Thr94)-H(WAT1)


constraint of 1.17 Å between H(PSB) and O1(Glu113)
distance (Å)
time (ps)
0
1
2
3
4
5
0
0.5
1
1.5
2
2.5
N(PSB)-H(PSB)
O1(Glu113)-H(PSB)
O2(Glu113)-Hbb(Cys187)
O1(Glu113)-H(Thr94)
O1(Glu113)-H(WAT1)
O(Thr94)-H(WAT1)

constraint of 1.07 Å between H(PSB) and O1(Glu113)



constraint of 1.00 Å between H(PSB) and O1(Glu113)

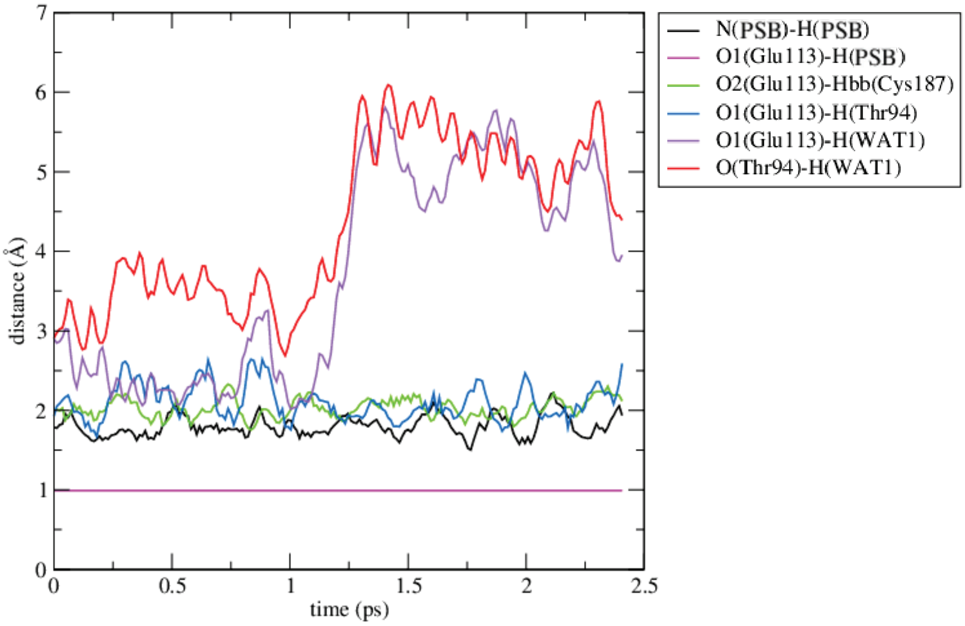


Table A.2: pKa calculations of Glu113 in different rhodopsin intermediates. In addition, the pKa values for Asp83 and Glu122 are shown as reference values since these residues are believed to be protonated all throughout the activation [Fahmy et al., 1993]. pKa calculations were performed with an external dielectric of 80, a salinity of 0.15 M and different internal dielectrics. The calculated shift is the total difference between the full (all) and the minimal (no Gly90, Thr94, WAT1, WAT2) model system, called *no G90, T94, waters*. Each time a molecule or residue is removed, the charges of the group are set to zero, but the radii are maintained. The full model includes rhodopsin without the chromophore and all water molecules present in the active site.

internal dielectric of 4									
	dark			**lumi**			**pre-meta I**		
conformation	E113	D83	E122	E113	D83	E122	E113	D83	E122
all	5.1	>12	>12	5.4	>12	>12	>12	>12	10.9
no G90	4.7	>12	>12	5.3	>12	>12	>12	>12	11.3
no T94	6.9	>12	>12	7.3	>12	>12	>12	>12	10.9
no WAT1	6.5	>12	>12	7.5	>12	>12	10.3	>12	11.9
no WAT1, WAT2	6.7	>12	>12	8.3	>12	>12	11.5	>12	11.6
no T94, waters	8.5	>12	>12	10.3	>12	>12	11	>12	11.8
no G90, T94, waters	8.0	>12	>12	10.1	>12	>12	10.0	>12	>12
shift	2.9	0.0	0.0	4.7	0.0	0.0	<-2.0	0.0	>1.1

internal dielectric of 6									
	dark			**lumi**			**pre-meta I**		
conformation	E113	D83	E122	E113	D83	E122	E113	D83	E122
all	4.9	>12	10.2	5.1	>12	10.0	10	>12	9.0
no G90	4.6	>12	10.3	5.0	>12	10.1	9.4	>12	9.2
no T94	6.1	>12	10.3	6.4	>12	10.1	9.7	>12	9.0
no WAT1	5.9	>12	10.2	6.5	>12	10.0	8.5	>12	9.5
no WAT1, WAT2	6	>12	10.2	7	>12	10.0	9.1	>12	9.3
no T94, waters	7.2	>12	10.2	8.4	>12	10.0	8.8	>12	9.4
no G90, T94, waters	6.9	>12	10.3	8.3	>12	10.1	8.2	>12	9.6
shift	2.0	0.0	0.1	3.2	0.0	0.1	-1.8	0.0	0.6

internal dielectric of 10									
	dark			**lumi**			**pre-meta I**		
conformation	E113	D83	E122	E113	D83	E122	E113	D83	E122
all	4.6	9.6	8.0	4.7	9.3	7.9	7.7	9.7	7.3
no G90	4.4	9.7	8.1	4.6	9.4	7.9	7.3	9.8	7.4
no T94	5.4	9.6	8.1	5.5	9.3	7.9	7.5	9.7	7.3
no WAT1	5.2	9.5	8.0	5.6	9.2	7.8	6.8	9.7	7.3
no WAT1, WAT2	5.4	9.5	8.0	6.0	9.2	7.9	7.2	9.6	7.4
no T94, waters	6.1	9.5	8.0	6.8	9.2	7.8	7.1	9.6	7.4
no G90, T94, waters	5.9	9.6	8.1	6.8	9.3	7.9	6.7	9.7	7.5
shift	1.3	0.0	0.1	2.1	0.0	0.0	-1.0	0.0	0.2

internal dielectric of 20									
	dark			**lumi**			**pre-meta I**		
conformation	E113	D83	E122	E113	D83	E122	E113	D83	E122
all	4.3	6.7	6.3	4.3	6.5	6.2	6.0	6.7	5.9
no G90	4.1	6.8	6.3	4.2	6.6	6.2	5.8	6.8	5.9
no T94	4.7	6.7	6.3	4.8	6.5	6.2	5.9	6.7	5.9
no WAT1	4.7	6.7	6.3	4.8	6.5	6.2	5.5	6.7	6.0
no WAT1, WAT2	4.7	6.7	6.3	5.1	6.5	6.2	5.7	6.7	5.9
no T94, waters	5.2	6.7	6.3	5.6	6.5	6.2	5.7	6.7	6.0
no G90, T94, waters	5.1	6.7	6.3	5.5	6.6	6.2	5.5	6.7	6.0
shift	0.8	0.0	0.0	1.2	0.1	0.0	-0.5	0.0	0.1

B Effect of N-Terminal Myristoylation on the Active Conformation of Gα_{i1}:GTP

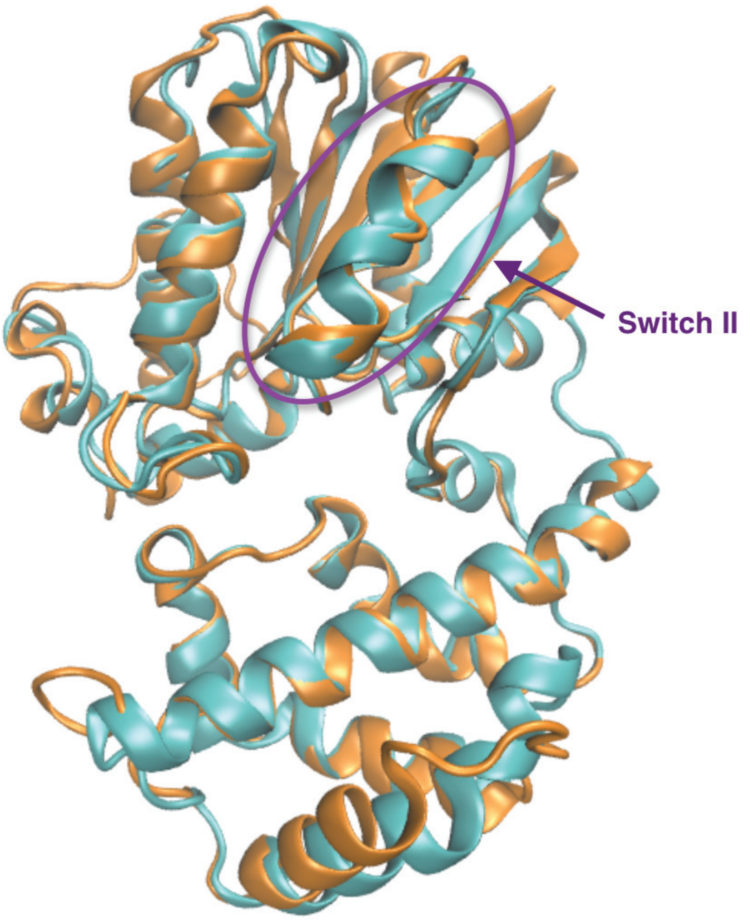


Figure B.1: Alignment of the Cα atoms of an active conformation of Gα_{i1}^{non}:GTPγS (cyan) and Gα_{s}^{non}:GTPγS (orange) taken from the PDB codes 1GIA and 1CJK. The root mean square deviation between both structures is 1.07 Å. GTPγS is a GTP analogue, which is not reactive in the active site of the subunit.

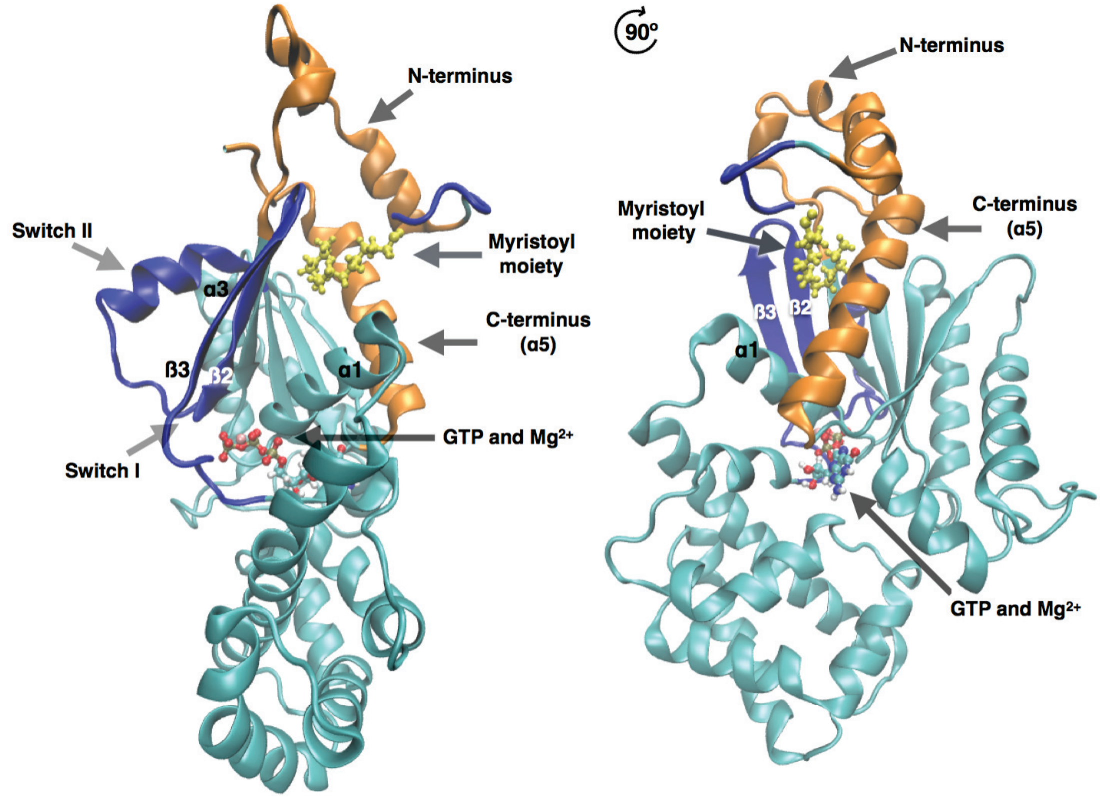
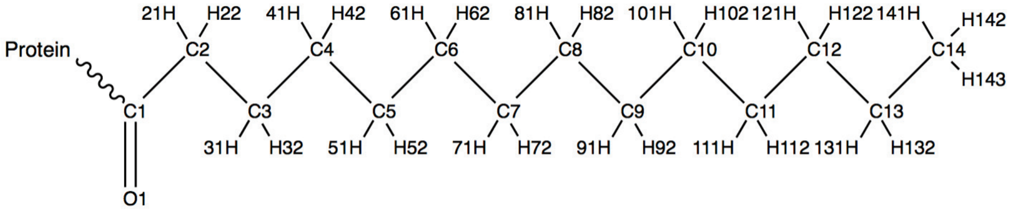


Figure B.2: Two perspectives of the initial homology model that was used to start the classical MD simulation. The template for the cyan region was PDB entry 4PAQ, for the orange area the PDB structure 1AS3 was used and the blue part was modelled after PDB entry 2K5U. The myristoyl moiety is depicted in yellow.

Figure B.3: Alignment of the Cα atoms of ARF1:GTP (PDB code 2KSQ) and Gα_{i1}^{non}:GDP (PDB code 1BOF) with an RMSD of 1.44 Å. The region on which ARF1:GTP is aligned with Gα_{i1}^{non}:GDP is shown in pink (residue 17-127) while the other regions are depicted in yellow. Gα_{i1}^{non}:GDP is represented in cyan.

Figure B.4: Myristoyl's atom names, atom types from the AMBER99SB force field and charges used for the generation of the moiety's topology.



Atom name	Atom type	Charge
C1	C	0.563313
O1	O	-0.558532
C2	CT	-0.268643
C3	CT	0.137672
C4	CT	-0.052827
C5	CT	-0.037691
C6	CT	-0.030556
C7	CT	0.061035
C8	CT	0.005799
C9	CT	-0.050847
C10	CT	-0.052887
C11	CT	0.077723
C12	CT	-0.163982
C13	CT	0.261740
C14	CT	-0.220513
H21	H1	0.080394
H22	H1	0.080394
H31	HC	-0.022591
H32	HC	-0.022591
H41	HC	0.035191
H42	HC	0.035191
H51	HC	0.011342
H52	HC	0.011342
H61	HC	0.009232
H62	HC	0.009232
H71	HC	-0.014679
H72	HC	-0.014679
H81	HC	-0.001793
H82	HC	-0.001793
H91	HC	0.013990
H92	HC	0.013990
H101	HC	0.020355
H102	HC	0.020355
H111	HC	-0.004362
H112	HC	-0.004362
H121	HC	0.026269
H122	HC	0.026269
H131	HC	-0.056392
H132	HC	-0.056392
H141	HC	0.045095
H142	HC	0.045095
H143	HC	0.045095

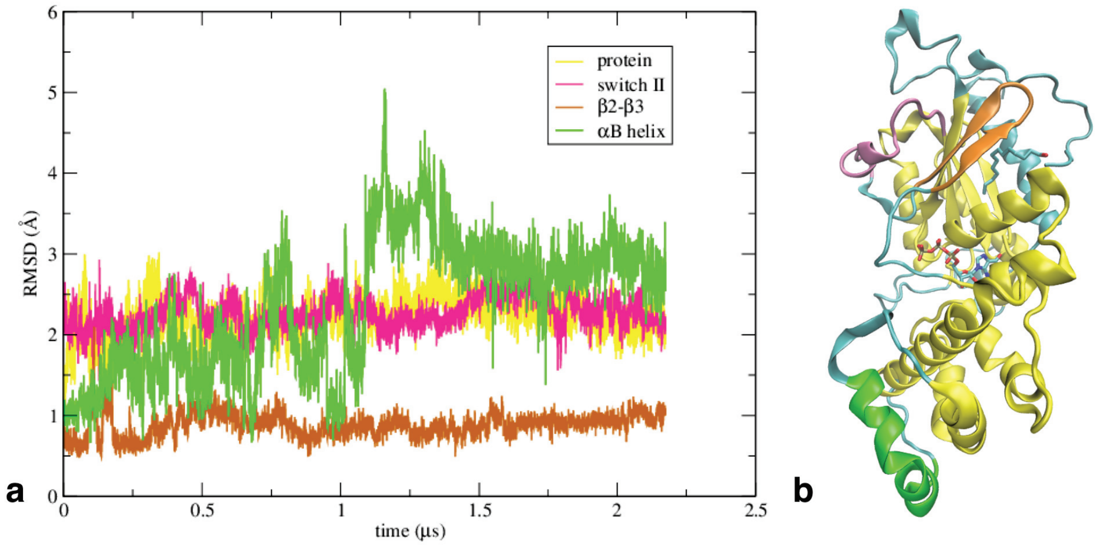


Figure B.5: Root mean square deviation of the backbone of several regions of the Gα_{i1}^{myr}:GTP complex. (a) Representation of the RMSD of the backbone of the protein in four regions: the switch II region (residues 206 to 220), the β2-β3 region (residues 187 to 199), the αB helix (residues 97 to 111) and the protein itself (residues 35 to 91, residues 121 to 178, residues 221 to 231, residues 242 to 279 and residues 296 to 327) for which the part of the protein that was not modelled was included and the αB helix was excluded together with highly flexible loops. (b) Representation of the regions that were used for the calculations of the RMSD. The colours of the regions overlap with the colours used in image a.

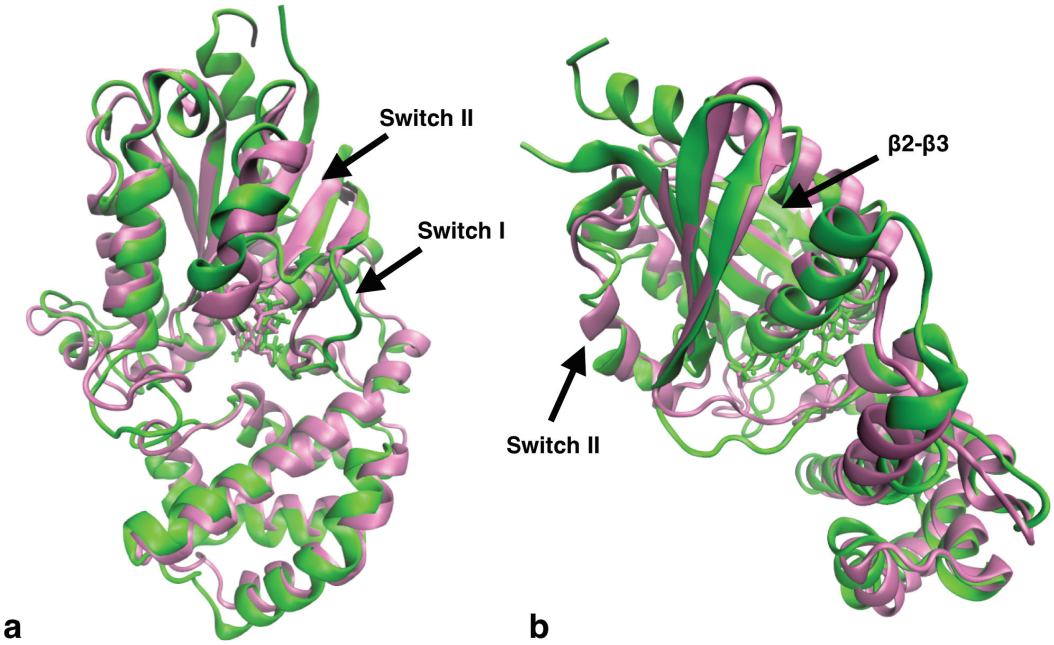


Figure B.6: Alignment of the Cα atoms of an X-ray of Gα_{i1}^{non} (PDB code 1GIA), interacting with Mg^{2+} and GTPγS, in mauve with the structure of the Gα_{i1}^{non}:GTP complex at ~1 μs of classical MD in green. (a) Orientation of the structural alignments in which the switch I and switch II regions can be viewed. Besides the protein, also the location of the GTP, GTPγS and Mg^{2+} are shown for both conformations. (b) Representation of the structural alignment in which the orientation of the β2-β3 region can be viewed. Besides the protein, also the location of the GTP, GTPγS and Mg^{2+} are shown for both conformations.

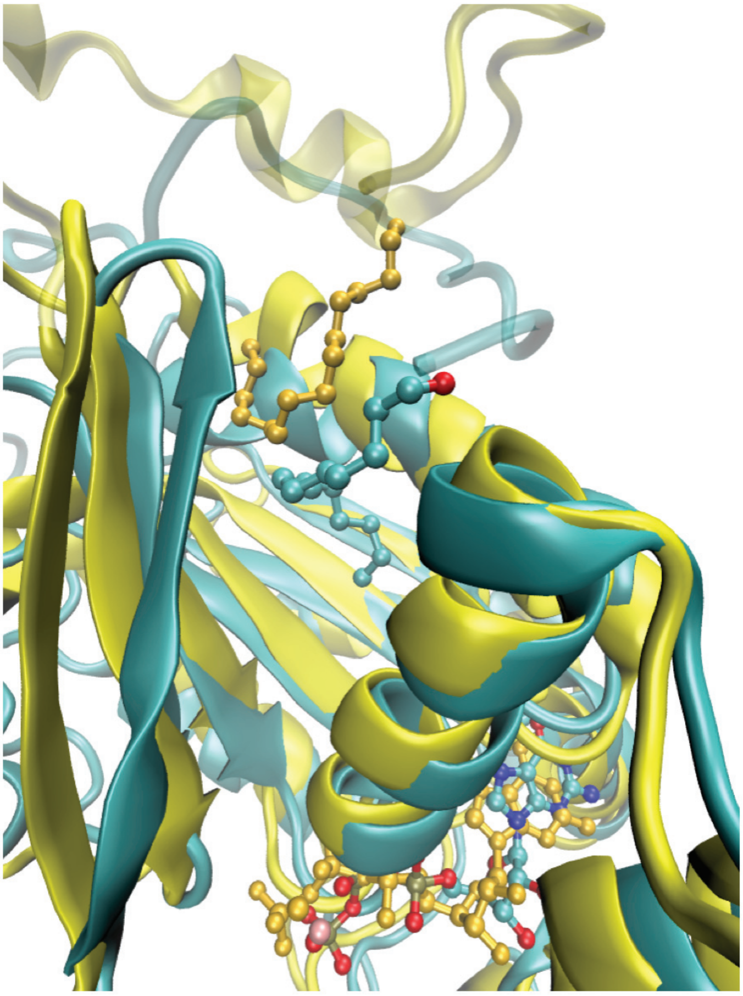


Figure B.7: Alignment of the Cα atoms of the initial conformation of the Gα_{i1}^{myr} model (yellow) with the conformation of Gα_{i1}^{myr} in the equilibrated conformation (cyan) at ~ 1.8 μs of classical MD. Besides the protein, also the location of the myristoyl group, GTP and Mg^{2+} are shown for both conformations.

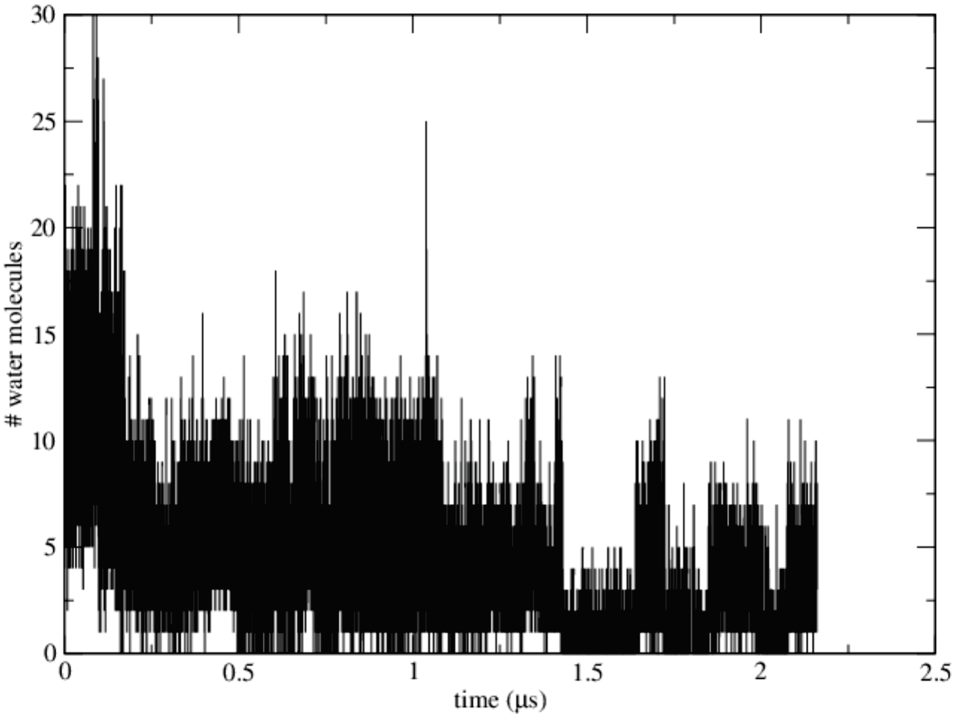


Figure B.8: Representation of the number of water molecules within 3 Å of the side chain of Trp211 during the classical MD simulation of Gα_{i1}^{myr}:GTP.

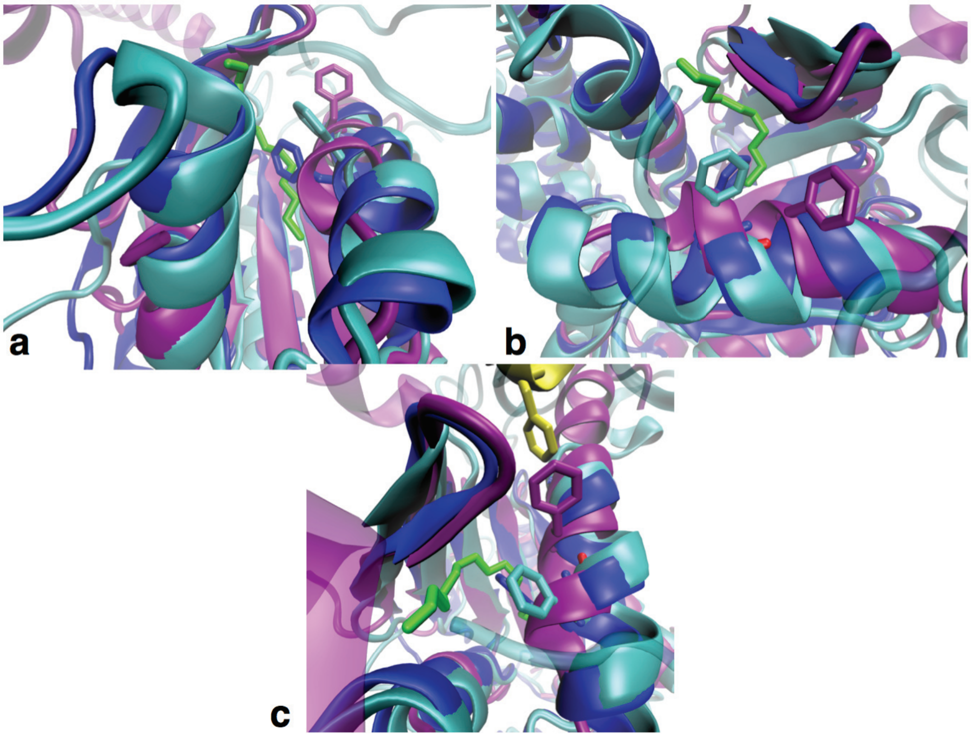


Figure B.9: Comparison of the orientation of Phe336 and Gα_s's equivalent Phe376 in different states of activation. (a) Aligned structures of Gα_{i1}^{myr}:GTP in cyan (after about 1.8 μs of classical MD), Gα_{i1}^{non}:GTPγS in blue (PDB code 1GIA) and Gs in purple. Gs is complexed to a GPCR, β2 adrenergic receptor, in yellow (PDB code 3SN6) in which Phe139's orientation is shown. The myristoyl moiety is shown in green. (b) Top view of the myristoyl binding pocket in which a clear shift of α5 helix can be observed for the Gs:GPCR complex. The colour scheme is the same as for image a. (c) Top view of the myristoyl binding pocket in which can be observed that the GPCR's Phe139 interacts with Gs's Phe376. The colour scheme is the same as for image a.

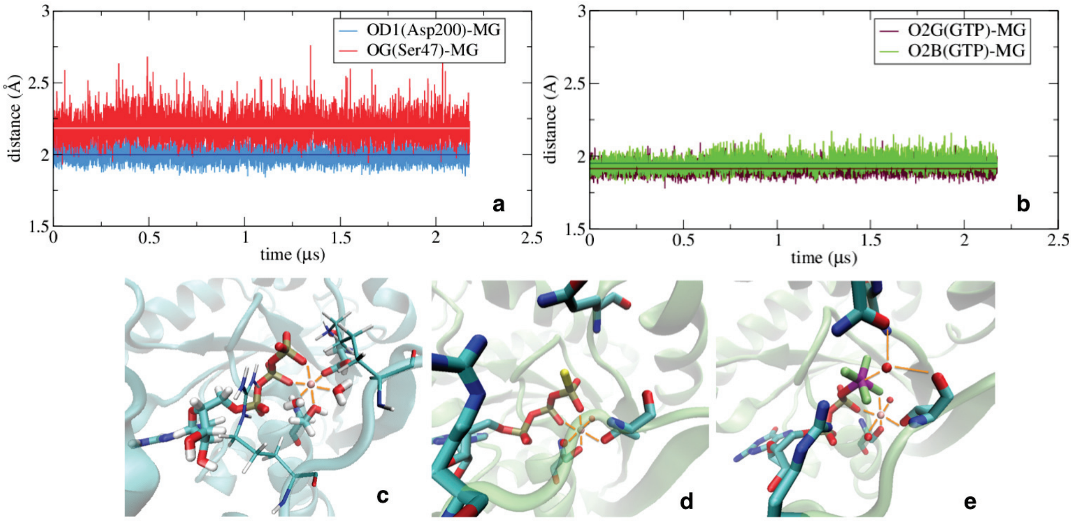


Figure B.10: Details of GTP binding in Gα_{i1}. (a) Distances between Mg^{2+} and two interacting ligands, Ser47 and Asp200, in the Gα_{i1}^{myr}:GTP complex. (b) Distances between Mg^{2+} and two interacting oxygen atoms from the GTP molecule in the Gα_{i1}^{myr}:GTP complex. (c) Zoom of the structure of the simulated Gα_{i1}^{myr}:GTP complex in which the GTP molecule and the Mg^{2+} are shown with relevant interacting residues. Characteristic hydrogen bonds are described by the orange lines. (d) Zoom on the structure of the Gα_{i1}^{non}:GTPγS complex (PDB code 1GIA) [Liu et al., 2009] in which the GTPγS molecule and the Mg^{2+} are shown with characteristic interacting residues: Ser47, Arg178, Thr181, Asp200, Gln204. Important hydrogen bonds between moieties are described by the orange lines. (e) Zoom on the structure of the Gα_{i1}^{non}:GDP:AlF_4^- complex (PDB code 1GFI) [Liu et al., 2009] in which the GDP molecule, the AlF_4^- ion and the Mg^{2+} ion are shown with relevant interacting residues: Ser47, Arg178, Thr181, Asp200, Gln204. Important hydrogen bonds are described by the orange lines.

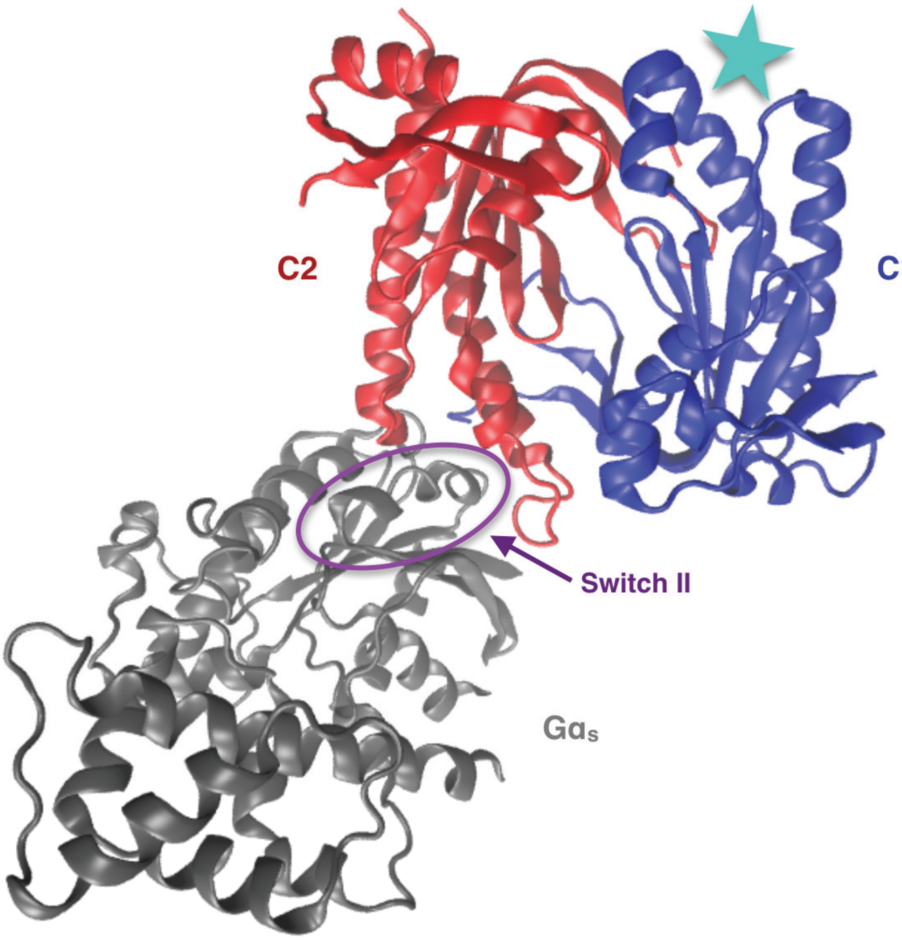


Figure B.11: View of the Gα_s:AC5 complex (PDB code 1CJK) from the cytosolic side. The Gα_s subunit is depicted in grey, while the C1 domain is represented in blue and the C2 domain is shown in red. The location of the Gα_{i1} structure is described by the cyan star.

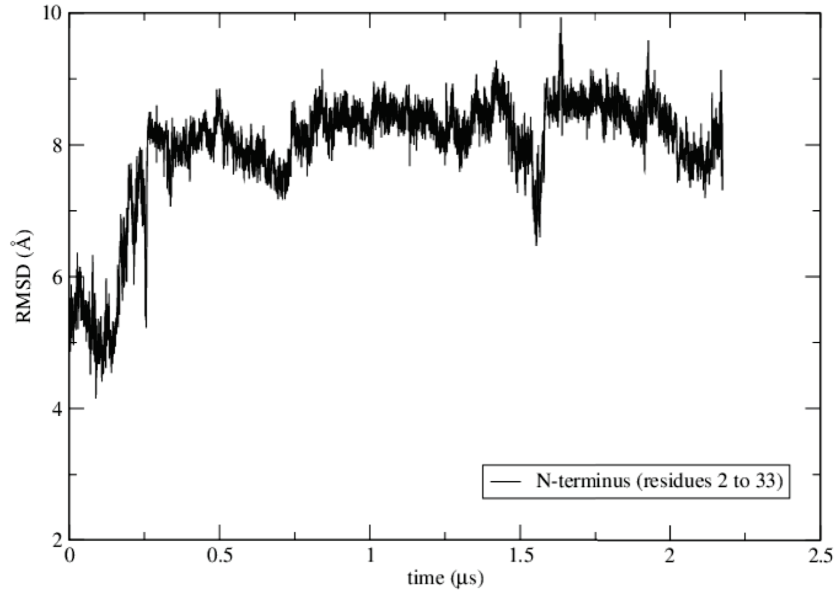


Figure B.12: Root mean square deviation of the backbone of the N-terminus of Gα_{i1}^{myr}:GTP over the entire classical MD trajectory.

B.1 Supporting Methods

One of the reasons for using ARF1 as a template for regions of Gα_{i1}^{myr} is that both, Gα_{i1}^{myr} and ARF1, contain a *Ras* domain. In addition, both proteins contain an N-terminal myristoyl group and are structurally very similar (Fig. 4.1), which is also shown via a root mean square deviation of 1.4 Å between ARF1:GTP and Gα_{i1}^{non}:GDP (Fig. B.3). Like Gα_{i1}^{myr}, ARF1 is also able to switch between its active and inactive state via GDP/GTP exchange for which NMR structures are available for both cases [Liu et al., 2009, 2010]. When GDP is interacting with an ARF1 protein, the protein is inactive and its conformation is an equilibrium between a solvated and a transiently membrane-bound state [Behnia and Munro, 2005]. In its solvated state, ARF1's myristoyl group is able to bury itself into an apolar binding site on the protein, while the N-terminal helix becomes unstructured but remains close to the protein [Liu et al., 2009]. Once GDP is exchanged for GTP in ARF1's nucleotide binding site, the protein will become active and is prone to bind to the cell membrane [Liu et al., 2010]. In the active state, the N-terminal helix maintains its structure and the myristoyl group is able to interact with the cell membrane as it has been removed from the protein binding pocket. Comparing the solvated ARF1:GDP and membrane-bound ARF1:GTP conformation, it becomes apparent that the switch II and the β2-β3 domains rearrange the most upon GDP/GTP exchange (Fig. 4.1). The domains have to adjust as an area is formed in which the myristoyl is able to dock. Therefore, the template of the ARF1:GDP complex was used to model this region of the soluble Gα_{i1}^{myr}:GTP complex. Even when the N-terminus is cut off and the myristoyl is absent, Δ17ARF1:GDP, the ARF1 complex is still able to form a hydrophobic pocket, which is similar to the myristoylated ARF1:GDP complex [Seidel et al., 2004]. Although the ARF1:GDP and Δ17ARF1:GDP structures are globally similar in their conformation, significant changes were observed by Seidel et al. around: the nucleotide binding site, the interswitch toggle region and the C-terminus [Seidel et al., 2004]. These are all regions that are in direct or indirect contact with the myristoyl moiety in the myristoylated ARF1:GDP complex. The formation of the hydrophobic pocket even in the Δ17ARF1:GDP complex could be due to the fact that ARF1 only contains a *Ras* domain. Therefore, the switch I region is connecting α1 and β3 directly, hence, when GTP is exchanged for GDP switch I is free to move and will adjust to a stable conformation inducing a large rearrangement in the ARF1 conformation (Fig. 4.1). In Gα_{i1}^{myr}'s conformation an alpha helical (AH) domain is present between α1 and the switch I region, which could significantly alter the possible conformational rearrangements that the protein can under go due to GDP/GTP exchange (Fig. 4.3).

B.2 Supporting Results and Discussion

B.2.1 Myristoyl Binding Site

The observed upward movement of Phe336 in the myristoylated structure of Gα_{i1}^{myr}, due to a change in the residue's rotamer conformation, is neither present in the Gα_{non}^{myr}:GTPγS nor in the inactive Gα_{i1}^{non}:GDP complex (Fig. 4.4a). Phe336's change of orientation is caused by the myristoyl group that is interacting with the apolar pocket at Phe336's initial position. The only X-ray structure in which similar behaviour of Phe336 has been observed is a G$_s$ structure that is in the nucleotide free state and is interacting with a GPCR, the β2 adrenergic receptor. This G$_s$ structure is on its way to the active conformation, but is captured in an intermediate state between active and inactive. The phenylalanine residue in G$_s$, Phe376 (UniprotKB P04896), that is equivalent to Gα_{i1}'s Phe336, moves away from the hydrophobic binding pocket because of its interactions with the present GPCR (Fig. B.9) [Rasmussen et al., 2011]. The β2 adrenergic receptor is known to be able to interact with the G$_s$ as well as the G$_{i1}$ protein [Kilts et al., 2000]. The effect of the GPCR on G$_s$'s conformation could therefore be similar to the effect the GPCR would have on the G$_i$ structure [Coleman et al., 1994]. Although similar reorientation for both the G$_s$:GPCR and the Gα_{i1}^{myr}:GTP complex can be observed, the α5 helix has moved and rotated towards the GPCR in the Gs:GPCR complex (Fig. B.9), which is not the case in the Gα_{i1}^{myr}:GTP structure as Phe336 reorients due to the presence of the myristoyl not because of a movement of the α5 helix. Hence, interesting similarities and differences can be observed between the aligned structures in figure B.8, which underline the importance of Phe336 in the activation process of G proteins.

Besides Phe336, the conformation of β2-β3 is also modified due to the presence of the myristoyl group (Fig. 4.4b). In figure 4.4 (a, b), the outward movement and shift of β2-β3 in the simulated Gα_{i1}^{myr}:GTP complex is depicted through an alignment with the Gα_{i1}^{non}:GDP (and Gα_{i1}^{non}:GTPγS) structures. Residues, such as Phe191 and Phe189 that are part of β2-β3 have to change their position due to the steric pressure induced by the myristoyl moiety (Fig. 4.6a, b). Because of this rearrangement in β2-β3 and the presence of the lipid, Phe191 and Phe336 are not able to interact with each other anymore like in the Gα_{i1}^{non}:GDP complex (Fig. 4.6b). The absence of direct interactions between Phe191 and Phe336 has been associated to an active state of the Gα_{i1} subunit [Kaya et al., 2014]. All these rearrangements in the myristoyl binding pocket are necessary to make space for the myristoyl moiety and prevent interactions between β2-β3 and α5 (Fig. 4.6a). The shift of β2-β3 with respect to the rest of the protein results in the addition of two residue units to the switch II region (Fig. 4.4b, c). Hence, in the Gα_{i1}^{myr}:GTP complex, the last hydrogen bond of β2-β3 between Asp200 and Val185 aligns with residues Gly202 and Gly183 in the Gα_{i1}:GDP complex.

The location of the myristoyl group in the simulated Gα_{i1}^{myr}:GTP complex stabilises in a similar hydrophobic binding pocket as in ARF1:GDP. However, Preininger *et al.* suggests another location, between α3 and the switch II region, for the myristoyl moiety on active Gα_{i1}^{myr} [Preininger et al., 2012]. This hypothesis is mainly based on two findings: a decrease in the solubility of Trp211 and a fluorescence quenching experiment to locate the N-terminus in the GTP-bound form.

Interestingly, the experimental data in combination with the newly obtained classical MD simulations can also be interpreted in a different way. Due to the conformational changes induced by the myristoyl group upon activation, the site around Trp211 becomes less solvent exposed in the simulation (Fig. B.8). The simulation also shows that due to the decrease in solvent accessibility, the initial space between α3 and the switch II region in Gα_{i1}^{non}:GTPγS is blocked by switch II (Fig. 4.3). This conformational change of switch II would significantly increase the difficulty of binding the myristoyl moiety between α3 and switch II. The quenching experiment performed on active Gα_{i1}^{myr} by Preininger et al. suggests that during Gα_{i1}^{myr}'s activation process the N-terminus decreases its distance with respect to the switch II region. Nonetheless, the region around switch II in which the N-terminus could be located, still covers a large region [Preininger et al., 2012] and leaves possibility open for other myristoyl binding sites, such as the hydrophobic binding pocket.

B.2.2 N-Terminus

In the initial structure of the classical MD simulations the X-ray structure of Gα_{i1}^{non}:GDP (PDB code 1AS3) was used as a template for the N-terminus. This X-ray structure was employed because no GTPγS bound Gα_{i1}^{non} structures contain the N-terminal region. The crystallisation conditions for 1AS3 include the presence of a SO_4^- ion, which could have an effect on the conformation of the N-terminus. However, experiments [Preininger et al., 2003] have shown that the conformation of the N-terminal region is mainly unstructured and flexible without the attachment of the myristoyl moiety, which would be similar to the X-ray structure of 1AS3 as in this system the myristoyl is not present. Therefore, the N-terminus of 1AS3 could be affected by the SO_4^- ion, but the experiments performed by Preininger *et al.* [Preininger et al., 2003] suggest that the non-myristoylated N-terminal region will lose its structural features upon solvation. When the myristoyl group is bound, the N-terminus becomes closely packed to the protein [Preininger et al., 2003]. In order for the N-terminus to decrease its distance with respect to the rest of the protein, the N-terminus has to change its orientation significantly compared to the heterotrimer conformation.

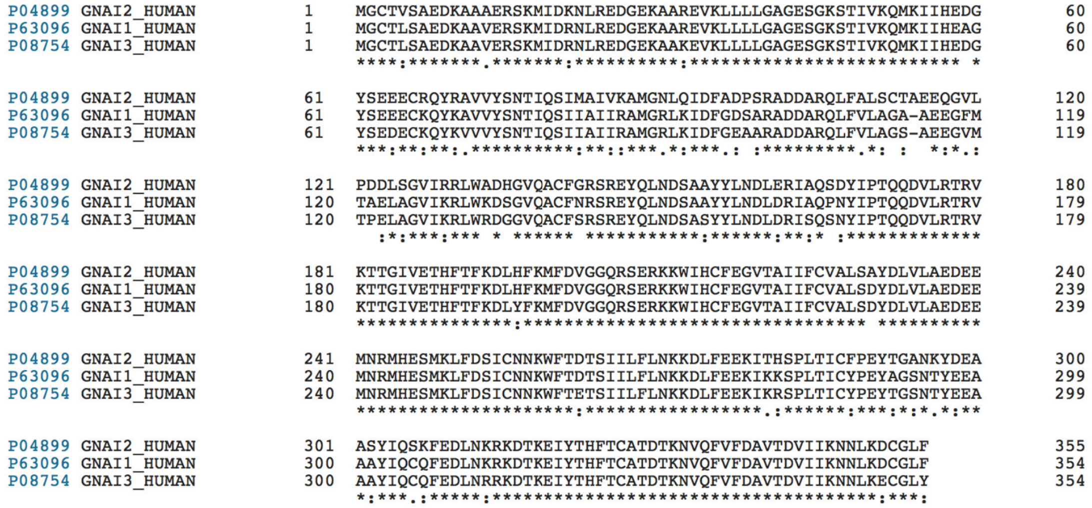
Hence, although the SO_4^- ion might affect the conformation of the X-ray structure, the N-terminus is bound to change its conformation significantly in the classical MD simulations as in the simulated system the myristoyl group is present. In the performed simulation the initially guessed N-terminal conformation indeed changes its shape significantly upon solvation (Fig. B.12), stabilising in a closely packed conformation on the protein (Fig. 4.3). At the beginning of the simulation, the myristoyl group is also located in a different position then its final location in the hydrophobic pocket which shows that it can sample other regions of the protein besides its final location (Fig. B.2).

B.2.3 GTP Binding Site

As a result of the slight shift of the β2-β3 residues, Asp200 which is located in this strand is able to interact with the Mg2+ ion in the Gα_{i1}^{myr}:GTP conformation (Fig. B.10). When comparing Gα_{i1}^{non}:GTPγS (PDB code 1GIA) to Gα_{i1}^{myr}:GTP, the Mg^{2+} ion interactions thus change from Ser47, Thr181, GTP and two water molecules, to Ser47, Asp200, GTP and two water molecules (Fig. B.10a, c). The water molecules that are interacting with the Mg^{2+} ion undergo water exchange during the classical MD trajectory. The triphosphate region of GTP is not only stabilised via Mg^{2+} interactions but also through other protein residues. These interactions change from the backbone of Glu43, Ser44, Lys46, Ser47, Thr48 in the Gα_{i1}^{non}:GTPγS conformation to the backbone of Glu43, Ser44, Lys46, Ser47, Thr48, Arg178 and Lys180 in the Gα_{i1}^{myr}:GTP complex. The newly formed interaction between GTP and Arg178, also known as the arginine finger, in the Gα_{i1}^{myr}:GTP complex is believed to be crucial for GTP hydrolysis [Ligeti et al., 2012]. The overall change of ligands coordinating to Mg^{2+} and GTP that occurs when moving from the Gα_{i1}^{myr}:GTPγS to the Gα_{i1}^{myr}:GTP structure, increases the polarity of the nucleotide binding site in the active myristoylated state, that is consistent with the fluorescence experiments of Preininger *et al.* in which the nucleotide binding site of myristoylated and non-myristoylated Gα_{i1}:GTPγS are compared [Preininger et al., 2012].

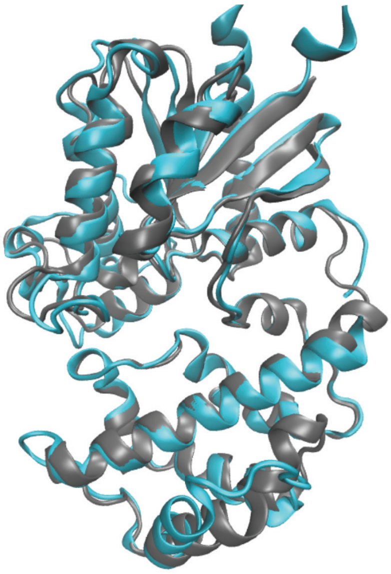
It is also interesting to compare the simulation results with an X-ray structure of the Gα_{i1}^{non} subunit that represents an intermediate state of Gα_{i1}:GTP during hydrolysis (PDB code 1GFI) (Fig. B.10e) [Coleman et al., 1994]. This intermediate conformation, however, has also been crystallised without the N-terminus and the myristoyl moiety. Also in this conformation, the Mg^{2+} ion interactions remain different from the one observed for the Gα_{i1}^{myr}:GTP complex. The solvated Gα_{i1}^{myr} subunit (Fig. B.10c) and the Gα_{i1}^{non}:GDP:AlF_4^- complex do share the same interaction between Mg^{2+} and Ser47 and Arg178's interactions with GTP or GDP:AlF_4^- (Fig. B.10e). However, the substitution of Thr181 by water and the exchange of a water molecule for Asp200 when going from the GDP:AlF_4^- complex to GTP are the same differences as the ones observed between the Gα_{i1}^{myr}:GTP and Gα_{i1}^{non}:GTPγS complexes. Another difference between the two complexes is the presence of a water molecule in the Gα_{i1}^{non}:GDP:AlF_4^- complex located between AlF_4^-, Gln204 and Thr181, which is absent in Gα_{i1}^{myr}:GTP. This water molecule in the Gα_{i1}^{non}:GDP:AlF_4^- complex that is proposed to be deprotonated in order to become the attacking base during GTP hydrolysis, is located in the position of Lys180 in the Gα_{i1}^{myr}:GTP conformation. Hence, in the case of the Gα_{i1}^{myr}:GTP complex, a conformational change and/or an additional protein interaction appears to be necessary in order to facilitate GTP hydrolysis. Even though a water molecule is present for GTP hydrolysis in the Gα_{i1}^{non}:GDP:AlF$_4^-$ conformation, an initial deprotonation step of the water molecule that would generate the attacking base seems unfavourable since the two hydrogens of the water molecule point towards Gln204 and the backbone of Thr181, which are not common proton acceptors. In conclusion, the interswitch toggle that influences the conformation of the switch II region also affects the GTP binding site significantly and leads to altered Mg^{2+} interactions impacting GTP hydrolysis.

C Exploring the Inhibition Mechanism of Adenylyl Cyclase Type 5 by N-terminal Myristoylated Gα_{i1}:GTP



```
P04899 GNAI2_HUMAN          1    MGCTVSAEDKAAAERSKMIDKNLREDGEKAAREVKLLLLGAGESGKSTIVKQMKIIHEDG      60
P63096 GNAI1_HUMAN          1    MGCTLSAEDKAAVERSKMIDRNLREDGEKAAREVKLLLLGAGESGKSTIVKQMKIIHEAG      60
P08754 GNAI3_HUMAN          1    MGCTLSAEDKAAVERSKMIDRNLREDGEKAAKEVKLLLLGAGESGKSTIVKQMKIIHEDG      60
                                 ****:*******.*******:**********:************************** *

P04899 GNAI2_HUMAN          61   YSEEECRQYRAVVYSNTIQSIMAIVKAMGNLQIDFADPSRADDARQLFALSCTAEEQGVL      120
P63096 GNAI1_HUMAN          61   YSEEECKQYKAVVYSNTIQSIIAIIRAMGRLKIDFGDSARADDARQLFVLAGA-AEEGFM      119
P08754 GNAI3_HUMAN          61   YSEDECKQYKVVVYSNTIQSIIAIIRAMGRLKIDFGEAARADDARQLFVLAGS-AEEGVM      119
                                 ***:**:**:.**********:**::***.*:***.: :*********.*: :  *:*.:

P04899 GNAI2_HUMAN          121  PDDLSGVIRRLWADHGVQACFGRSREYQLNDSAAYYLNDLERIAQSDYIPTQQDVLRTRV      180
P63096 GNAI1_HUMAN          120  TAELAGVIKRLWKDSGVQACFNRSREYQLNDSAAYYLNDLDRIAQPNYIPTQQDVLRTRV      179
P08754 GNAI3_HUMAN          120  TPELAGVIKRLWRDGGVQACFSRSREYQLNDSASYYLNDLDRISQSNYIPTQQDVLRTRV      179
                                   :*:***:*** * ****** ***********:******:**:* :*************

P04899 GNAI2_HUMAN          181  KTTGIVETHFTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSAYDLVLAEDEE      240
P63096 GNAI1_HUMAN          180  KTTGIVETHFTFKDLHFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLVLAEDEE      239
P08754 GNAI3_HUMAN          180  KTTGIVETHFTFKDLYFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLVLAEDEE      239
                                 ***************:********************************* **********

P04899 GNAI2_HUMAN          241  MNRMHESMKLFDSICNNKWFTDTSIILFLNKKDLFEEKITHSPLTICFPEYTGANKYDEA      300
P63096 GNAI1_HUMAN          240  MNRMHESMKLFDSICNNKWFTDTSIILFLNKKDLFEEKIKKSPLTICYPEYAGSNTYEEA      299
P08754 GNAI3_HUMAN          240  MNRMHESMKLFDSICNNKWFTETSIILFLNKKDLFEEKIKRSPLTICYPEYTGSNTYEEA      299
                                 *********************:*****************.:******:***:*:*.*:**

P04899 GNAI2_HUMAN          301  ASYIQSKFEDLNKRKDTKEIYTHFTCATDTKNVQFVFDAVTDVIIKNNLKDCGLF           355
P63096 GNAI1_HUMAN          300  AAYIQCQFEDLNKRKDTKEIYTHFTCATDTKNVQFVFDAVTDVIIKNNLKDCGLF           354
P08754 GNAI3_HUMAN          300  AAYIQCQFEDLNRRKDTKEIYTHFTCATDTKNVQFVFDAVTDVIIKNNLKECGLY           354
                                 *:***.:*****:*************************************:***:
```

a



b

Figure C.1: Similarities in sequence and structure of Gα_i and Gα_s subunits. (a) Sequence alignment of human Gα_{i1}, Gα_{i2}, Gα_{i3}, which shows a 84% sequence identity between the subunits. (b) Structural alignment of GTP-analogue-bound Gα_{i1} (PDB code 1AZT) in cyan and GTP-analogue-bound Gα_s (PDB code 1AS0) in grey.

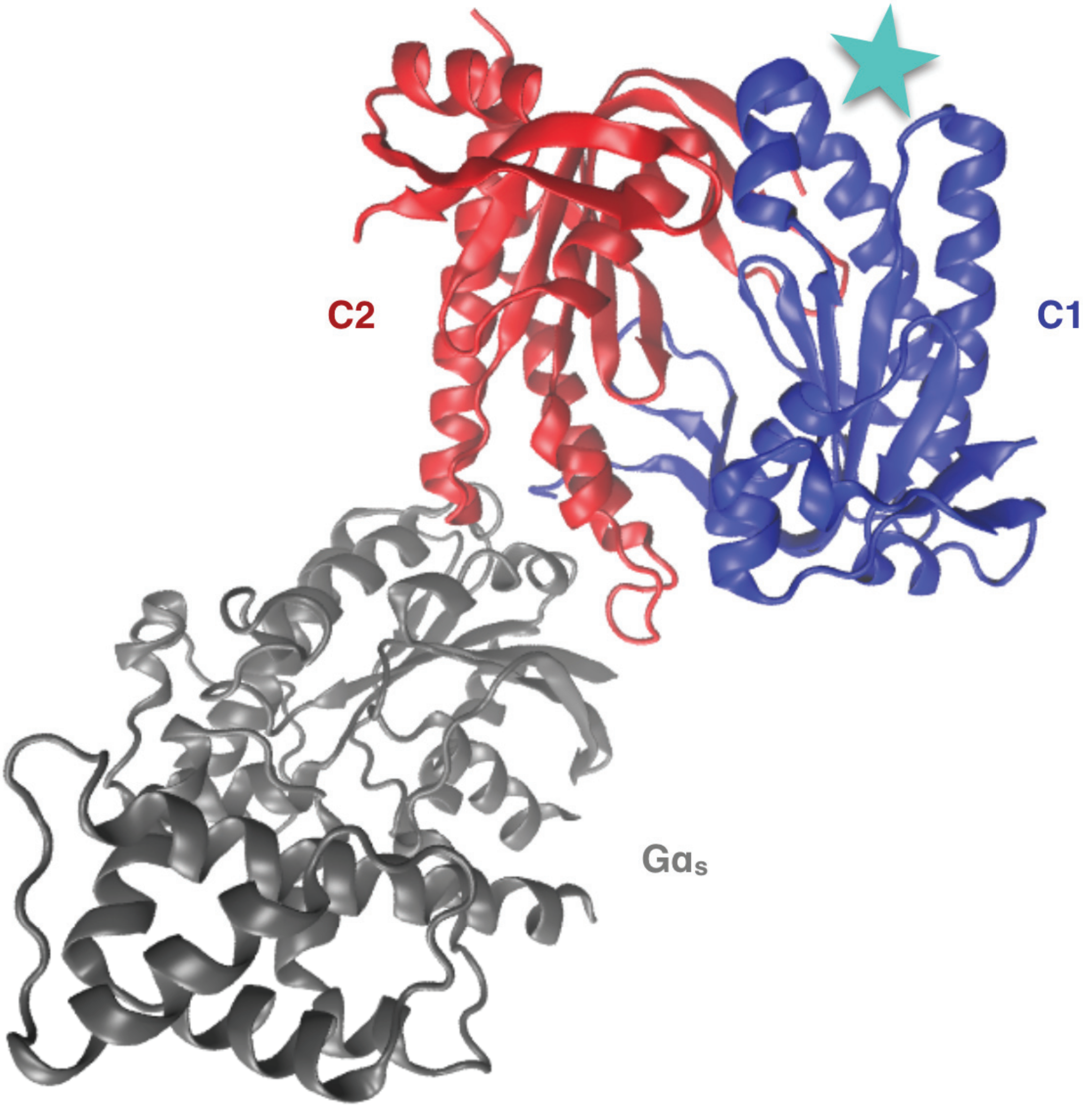


Figure C.2: View of the Gα_s:AC5 complex (PDB code 1CJK) from the cytosolic side. The Gα_s subunit is depicted in grey, while the C1 domain is represented in blue and the C2 domain is shown in red. The location of the Gα_{i1} structure is described by the cyan star.

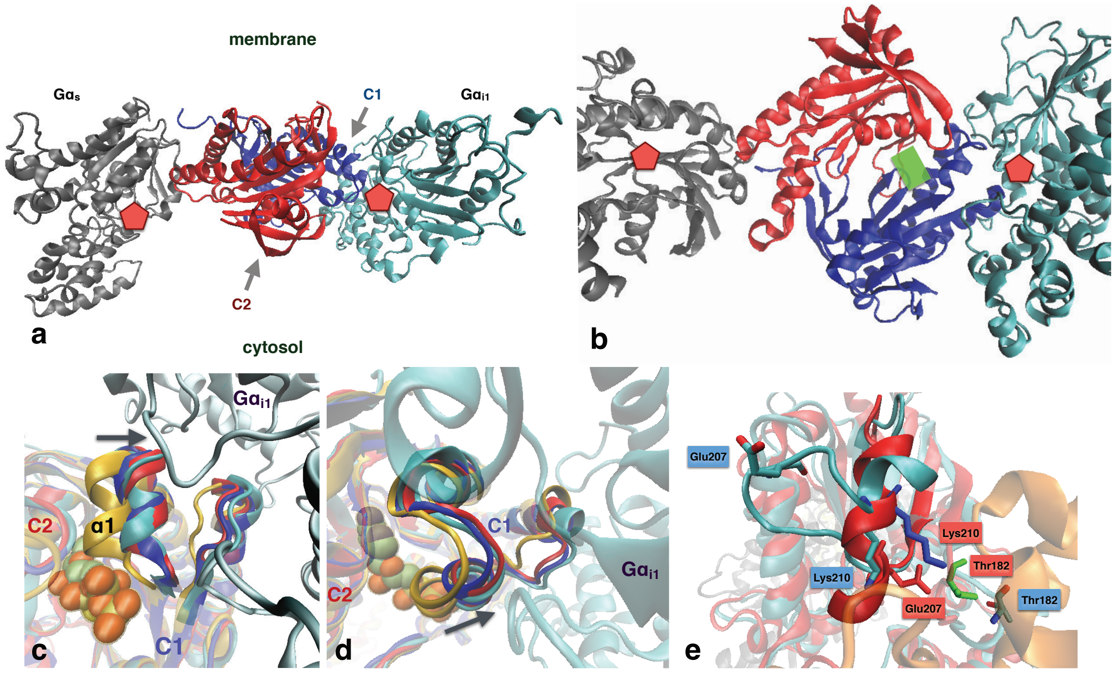


Figure C.3: Gα_{i1}^{myr}:GTP interacting with AC type 5 shows a different type of interaction mode with AC5 than Gα_s. (a) Representation of the docked Gα_{i1}^{myr} : $AC5$ complex of Rattus norvegicus with the location of Gα_s depicted as well to show the difference in association between the two Gα subunits. Gα_s is depicted in *gray*, Gα_{i1}^{myr} in *cyan*, the C1 domain in *blue* and the C2 domain in *red*. The location of the GTP molecules in both Gα subunits is represented by the *red* pentagons. (b) View from the cytosolic side on the docked Gα_{i1}^{myr}:AC5 complex, showing the position of the ATP molecule in the catalytic C1 domain in green. The color scheme is the same as in (a). (c, d) The difference in active AC conformation is shown, depending on AC's interaction partners (e.g. analogue of the substrate ATP, the inhibitor Ca^{2+}, substrate free). The alignment of the initial Gα_{i1}^{myr}:AC5 complex with three AC:Gα_s complexes demonstrates that the initial AC5 structure used in this study, the substrate-free state, is different from the fully active AC conformation (yellow) to which an ATP analogue is bound. The *yellow* structure is a complex of AC:Gα_s with an ATP analogue and forskolin (PDB code 1CJK). The ATP analogue is depicted in *yellow* as well with oxygen atoms in red, phosphor in tan and carbon in cyan. This structure is an active conformation of the AC catalytic domain. The *blue* (PDB code 3MAA) [Mou et al., 2009] is suggested to be an inactive AC:Gα_s complex and interacts with methylpiperazinoforskolin (FKP) together with an ATP analogue and a Ca^{2+} ion. The *red* structure (PDB code 1AZS) is an Gα_s:AC complex that only interacts with FKP in the catalytic domain and is more similar to the 3MAA structure than the fully active 1CJK structure around AC's active site. (e) Alignment of the Gα_{i1}^{non}:RGS4 complex (PDB code 1AGR) and the myristoylated Gα_{i1}:GTP complex. In case of the non-myristoylated Gα_{i1}:RGS4 structure (Gα_{i1}^{non}:RGS4), Gα_{i1}^{non} is shown in *red* and RGS4 is shown in *orange*. The myristoylated Gα_{i1} complex is depicted in *cyan*. The location of Thr182, Glu207 and Lys210 are shown for both complexes as these residues are important for the interaction between RGS4 and Gα_{i1} [Tesmer et al., 1997]. The Gα_{i1}^{non}:RGS4 residues are labeled in *red* and the Gα_{i1}^{myr}:GTP residues are labeled in *blue*.

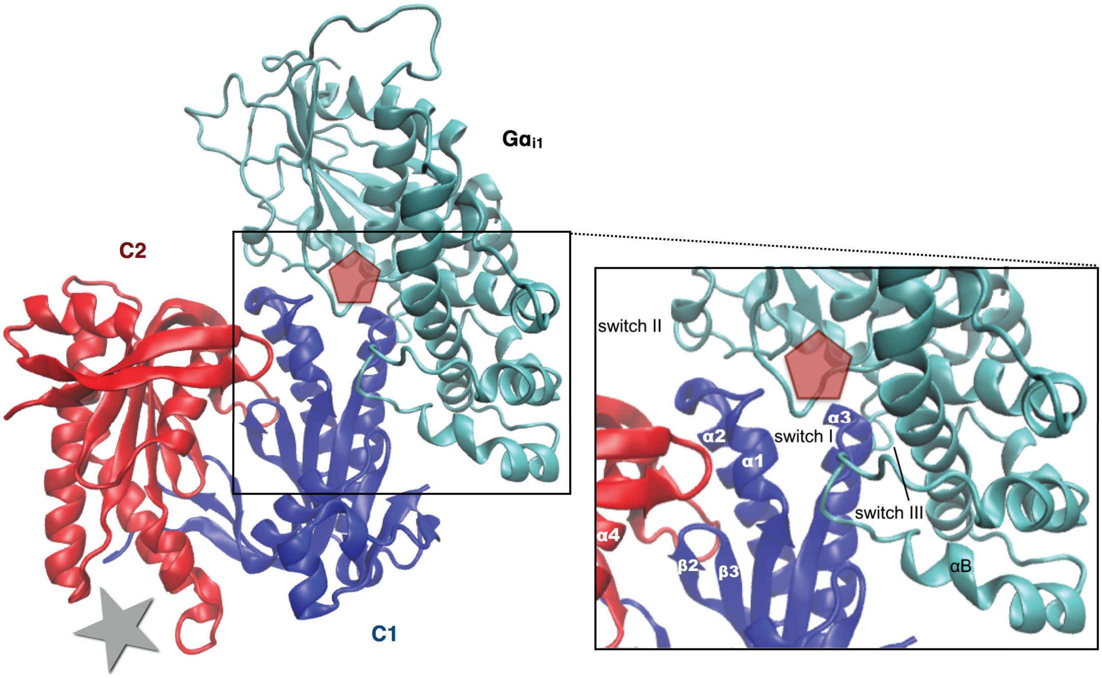


Figure C.4: View of the docked Gα_{i1}^{myr}:AC5 complex from the cytosolic side. The Gα_{i1}^{myr} subunit is depicted in cyan, while the C1 domain is represented in blue and the C2 domain is shown in red. The location of the Gα_s structure is described by the grey star and the GTP molecule is represented by the red pentagon.

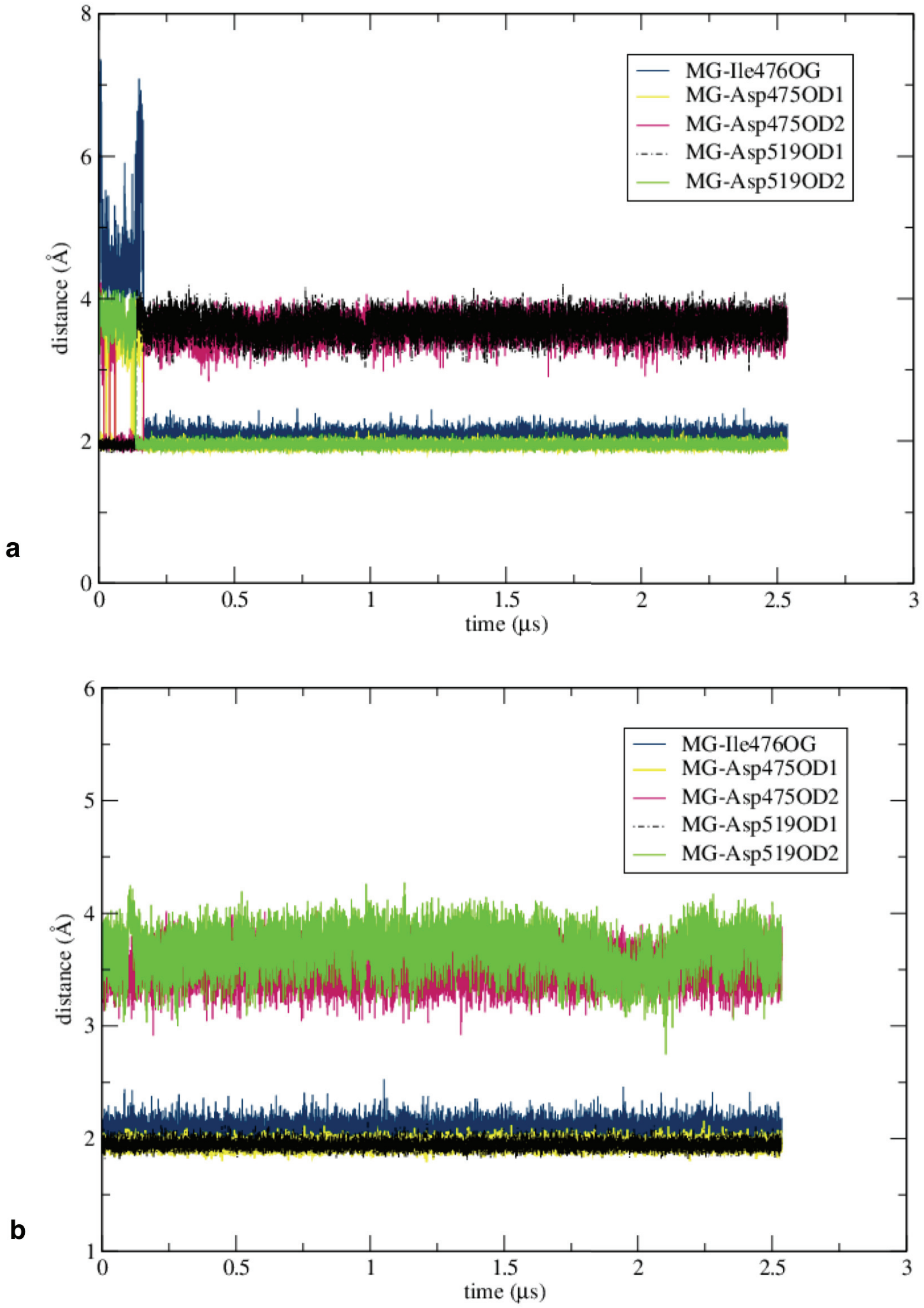


Figure C.5: Distances between the Mg^{2+} ion and residues in the active site of AC. (a) Distances between Mg^{2+} and residues in the active site of the $G\alpha_{i1}^{myr}$:AC5 system. (b) Distances between Mg^{2+} and residues in the active site of the free AC5 system.

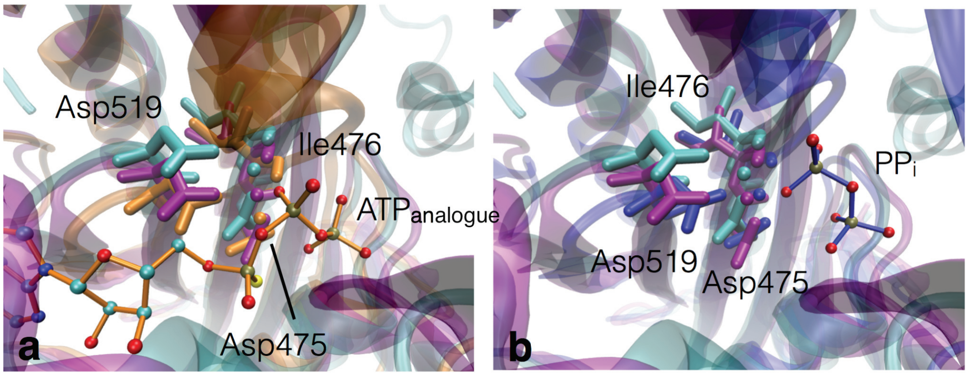


Figure C.6: Mg^{2+} ion in the active site of AC of the $G\alpha_{i1}^{myr}$:AC5 complex is located in the same interaction position as in ATP or pyrophosphate bound X-ray structures of $G\alpha_s$:AC complexes. (a) Alignment of the $G\alpha_{i1}^{myr}$:AC5 complex (cyan), the AC system (purple) and activated AC (PDB code 1CJK) in orange, which is interacting with ATP, showing the active site of AC at the C1/C2 interface. The residue names are following the Rattus norvegicus numbering for AC5. (b) Alignment of the $G\alpha_{i1}^{myr}$:AC5 complex (cyan), the AC system (purple) and AC associated to pyrophosphate, PPi, (PDB code 3C15) in blue showing the active site of AC at the C1/C2 interface. The residue names are following the Rattus norvegicus numbering for AC5.

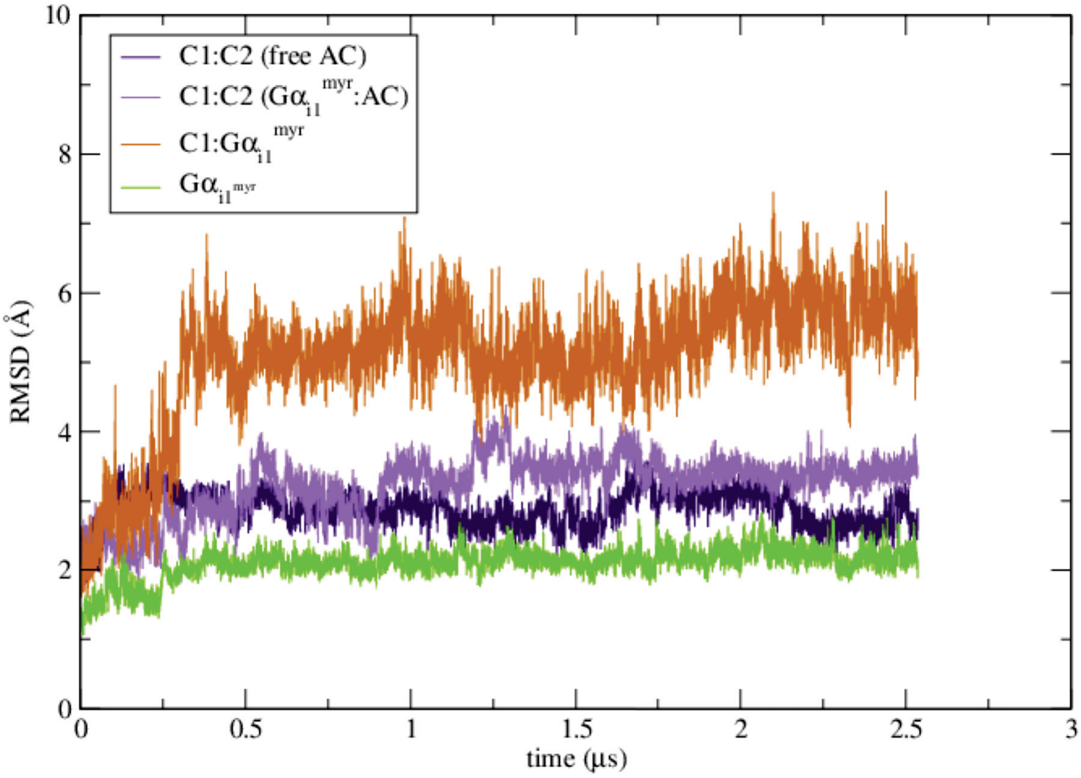


Figure C.7: Root mean square deviations of the backbone of the C1:C2 dimer. Additionally also the RMSD of the Gα_{i1}^{myr} subunit in the Gα_{i1}^{myr}:AC5 is shown, together with the RMSD of the combination of Gα_{i1}^{myr} and the C1 domain. In the RMSD calculation the residues between 463 to 644 were taken into account for the C1 domain, residues between 1065 to 1135 and 1145 to 1257 were used for the C2 domain and residues 34 to 334 were included for the Gα_{i1}^{myr} subunit.

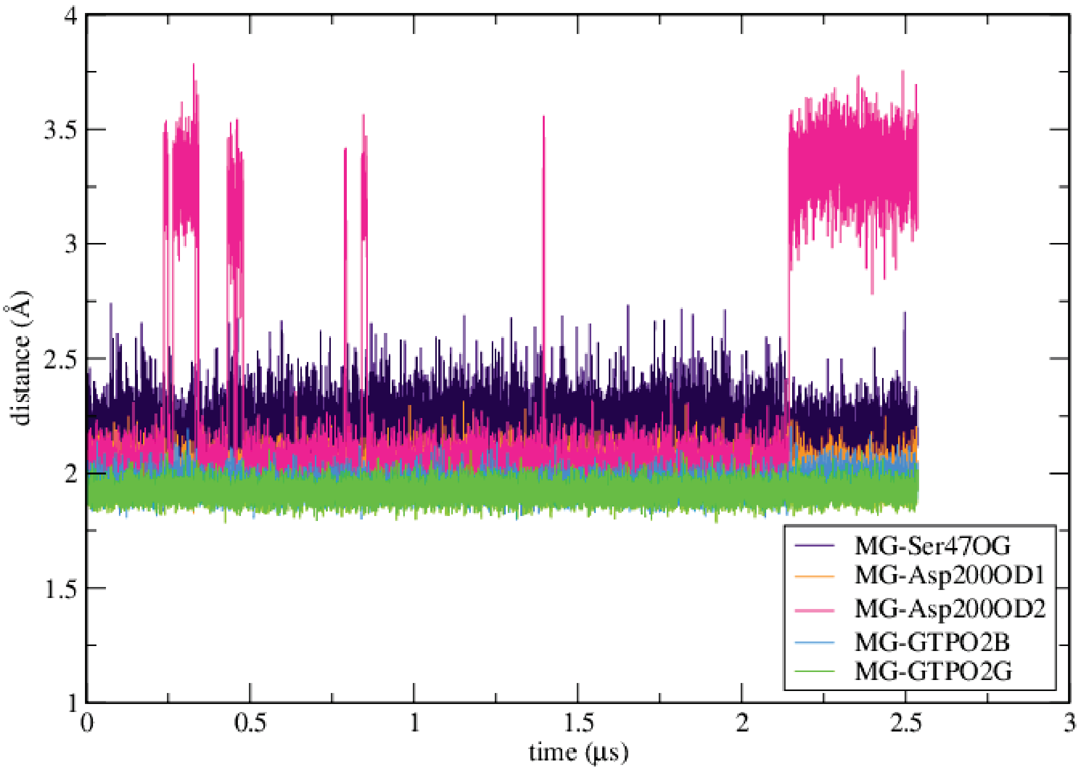


Figure C.8: Graph of the distances between the Mg^{2+} ion in the active site of $G\alpha_{i1}^{myr}$. Distances are shown between the Mg^{2+} ion and its environment, including GTP and the $G\alpha_{i1}^{myr}$ residues that are coordinating to Mg^{2+}: Ser47, Asp200.

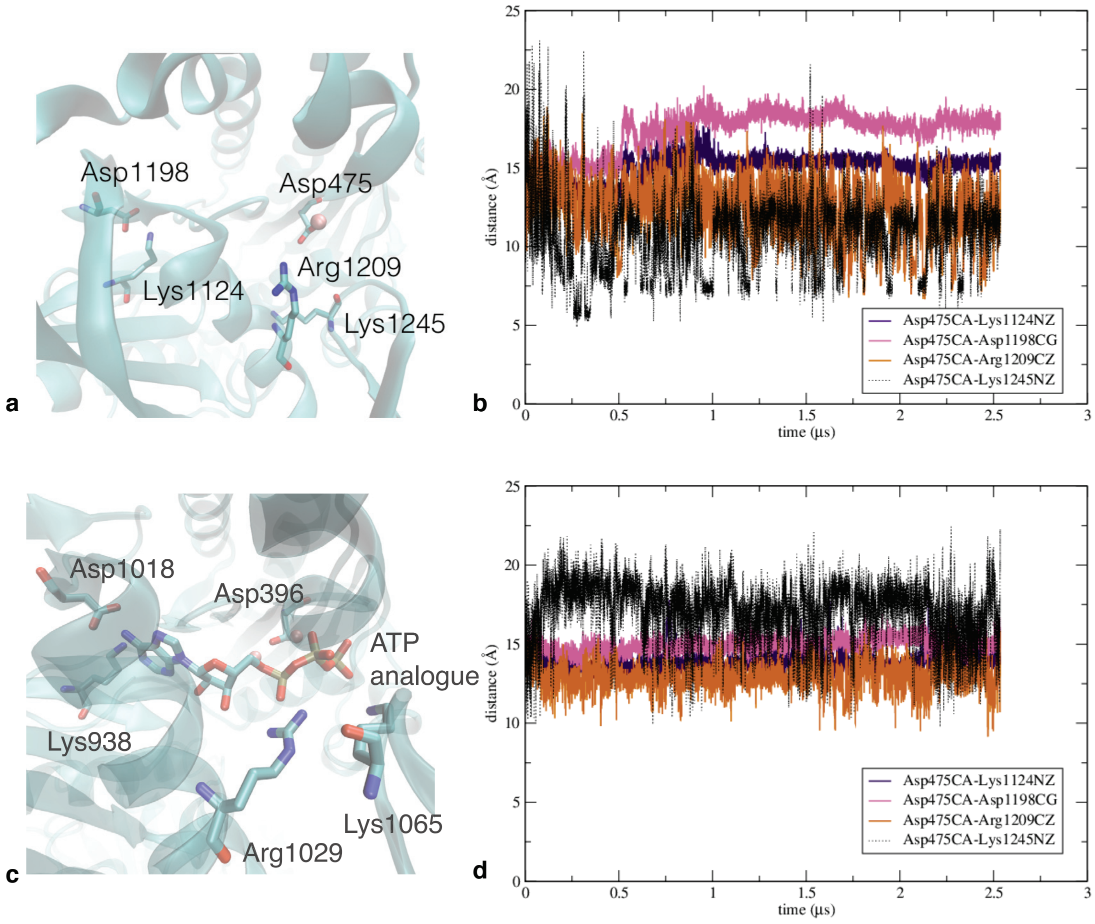


Figure C.9: Changes in the location of important residues in the active site of AC. (a) Detail of the active site of AC in the Gα_{i1}^{myr}:AC5 complex, showing the residues that are used in the distance calculations for image b and d. Additionally, the position of the Mg^{2+} ion is shown in pink. (b) Graph of the distances in the Gα_{i1}^{myr}:AC5 system between the Cα carbon of Asp475, which is positioned close to the ATP binding site, and other important residues for ATP conversion: Lys1124, Asp1198, Arg1209 and Lys1245. (c) Detail of the active site of AC in the Gα_s:AC complex in which an analogue of ATP is bound (PDB code 1CJK), showing the equivalent residues of the residues in AC5 that are used in the distance calculations for image b and d. (d) Graph of the distances in the AC system between the Cα carbon of Asp475, which is positioned close to the ATP binding site, and other important residues for ATP conversion: Lys1124, Asp1198, Arg1209 and Lys1245.

C.1 Supporting Results and Discussion

C.1.1 Docking of $G\alpha_{i1}^{myr}$ to AC5

$G\alpha_{i1}^{myr}$:AC5 Complex Suggests a Tighter Binding Mode of $G\alpha_{i1}^{myr}$ on AC Compared to $G\alpha_s$

Figure C.3a and C.3b depict the conformation of the initial $G\alpha_{i1}^{myr}$:AC5 structure, used in the classical MD simulation. In order to show the difference in orientation between $G\alpha_{i1}^{myr}$ on C1 and $G\alpha_s$ on C2, $G\alpha_s$'s location (as observed in the crystal structure) is also depicted in the figure. The selected complex from the docking results is characterised by high surface overlap with the C1 domain and a non-existing overlap with the C2 domain (Fig. C.4), which is consistent with the experimental data of Dessauer *et al.* [Dessauer et al., 1998]. $G\alpha_{i1}^{myr}$ also covers the experimentally proposed pseudo-symmetric $G\alpha_{i1}$:AC binding site on the C1 domain of AC5, the cleft formed by helices $\alpha 2$ and $\alpha 3$ [Dessauer et al., 1998] (Fig. C.4). Yet, unlike the suggested experimental mode of interaction, $G\alpha_{i1}^{myr}$ does not bind within the $\alpha 2$-$\alpha 3$ cleft, but clamps around the helices, forming a tight mode of interaction with AC5 (Fig. C.2 and Fig. C.4).

An additional validation of the proposed model of AC5 and $G\alpha_{i1}^{myr}$ is the comparison of active $G\alpha_{i1}^{myr}$ and the complex of non-myristoylated $G\alpha_{i1}$ ($G\alpha_{i1}^{non}$) and RGS4 as it has been suggested that the $G\alpha_{i1}^{myr}$:AC complex is insensitive to RGS4, a regulator of G-protein signalling (Fig. C.3e) [Markus et al., 2013]. When comparing the $G\alpha_{i1}^{myr}/$AC and the $G\alpha_{i1}^{non}/$RGS4 interface (PDB code 1AGR) [Tesmer et al., 1997], it is clear that residues in the switch II and switch I region of $G\alpha_{i1}^{myr}$, such as Thr182 and Glu210, that are important for RGS4 binding, are shifted compared to $G\alpha_{i1}^{non}$ in the RGS4 complex (Fig. C.3,e). Since AC5's C1 domain is interacting with and therefore restraining the conformation of $G\alpha_{i1}^{myr}$:GTP, the solvated $G\alpha_{i1}^{myr}$ appears to be incompatible with RGS4 binding due to the shift in the switch I and switch II region. This shift could be a reason for $G\alpha_{i1}^{myr}$'s insensitivity to RGS4 when complexed to AC.

C.1.2 Addition of Mg^{2+} to AC's Active Site Produces a Conformation Close to ATP or Pyrophosphate Bound Sites

The initial location of the Mg^{2+} ion was chosen close to the known Mg^{2+} sites to study whether the ion would migrate to one of the locations known via X-ray crystallography. Indeed a relocation of the Mg^{2+} ion occurs in both systems (free AC and the $G\alpha_{i1}^{myr}$:AC complex) into the site in which a Mg^{2+} ion is present in the ATP analogue and the pyrophosphate bound form (Fig. C.5 and Fig. C.6). This conformation of AC was deemed appropriate for the simulations as it represents one of the active conformations $G\alpha_{i1}^{myr}$ could be able to deactivate.

Bibliography

Allnér, O., Nilsson, L., and Villa, A. Magnesium ion–water coordination and exchange in biomolecular simulations. *Journal of Chemical Theory and Computation*, 8(4):1493–1502, 2012.

Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. H++ 3.0: automating pk prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*, 40(W1):W537–W541, 2012.

Atkins, P. W. and Friedman, R. S. *Molecular quantum mechanics*, pages 249–253. Oxford University Press, 2004a.

Atkins, P. W. and Friedman, R. S. *Molecular quantum mechanics*, pages 316–336. Oxford University Press, 2004b.

Atkins, P. W. and Friedman, R. S. *Molecular quantum mechanics*, pages 187–188. Oxford University Press, 2004c.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18):10037–10041, 2001.

Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.

Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6):3098, 1988.

Behnia, R. and Munro, S. Organelle identity and the signposts for membrane traffic. *Nature*, 438(7068):597–604, 2005.

Bekker, H., Berendsen, H. J. C., Dijkstra, E. J., Achterop, S., Van Drunen, R., Van der Spoel, D., Sijbers, A., Keegstra, H., Reitsma, B., and Renardus, M. K. R. Gromacs: A parallel computer for molecular dynamics simulations. *Physics Computing*, 92:252–256, 1993.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

Borea, P. A., Varani, K., Vincenzi, F., Baraldi, P. G., Tabrizi, M. A., Merighi, S., and Gessi, S. The a3 adenosine receptor: history and perspectives. *Pharmacological Reviews*, 67(1):74–102, 2015.

Bulo, R. E., Ensing, B., Sikkema, J., and Visscher, L. Toward a practical method for adaptive qm/mm simulations. *Journal of Chemical Theory and Computation*, 5(9):2212–2221, 2009.

Cabrera-Vera, T. M., Vanhauwe, J., Thomas, T. O., Medkova, M., Preininger, A., Mazzoni, M. R., and Hamm, H. E. Insights into g protein structure, function, and regulation. *Endocrine Reviews*, 24(6):765–781, 2003.

Campomanes, P., Neri, M., Horta, B. A. C., Röhrig, U. F., Vanni, S., Tavernelli, I., and Rothlisberger, U. Origin of the spectral shifts among the early intermediates of the rhodopsin photocycle. *Journal of the American Chemical Society*, 136(10):3842–3851, 2014.

Carloni, P., Rothlisberger, U., and Parrinello, M. The role and perspective of ab initio molecular dynamics in the study of biological systems. *Accounts of Chemical Research*, 35(6):455–464, 2002.

Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham III, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., and Goetz, A. W. Amber 2015; university of california: San francisco, ca, 2015. *There is no corresponding record for this reference*, 2015.

Ceperley, D. M. and Alder, B. J. The calculation of the properties of metallic hydrogen using monte carlo. *Physica B+ C*, 108(1-3):875–876, 1981.

Chen, C. A. and Manning, D. R. Regulation of g proteins by covalent modification. *Oncogene*, 20(13):1643–1652, 2001.

Cohen, G. B., Oprian, D. D., and Robinson, P. R. Mechanism of activation and inactivation of opsin: role of glu113 and lys296. *Biochemistry*, 31(50):12592–12601, 1992.

Coleman, D. E., Berghuis, A. M., Lee, E., Linder, M. E., and Gilman, A. G. Structures of active conformations of gi alpha 1 and the mechanism of gtp hydrolysis. *Science*, 265(5177): 1405–1412, 1994.

Consortium, U. Uniprot: a hub for protein information. *Nucleic Acids Research*, page gku989, 2014.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

CPMD. cpmd package. URL http://www.cpmd.org/. Copyright IBM Corp 1990-2015, Copyright MPI für Festkörperforschung Stuttgart 1997-2001.

Crocker, E., Eilers, M., Ahuja, S., Hornak, V., Hirshfeld, A., Sheves, M., and Smith, S. O. Location of trp265 in metarhodopsin ii: implications for the activation mechanism of the visual receptor rhodopsin. *Journal of Molecular Biology*, 357(1):163–172, 2006.

De Vries, S. J., van Dijk, M., and Bonvin, A. M. J. J. The haddock web server for data-driven biomolecular docking. *Nature Protocols*, 5(5):883–897, 2010.

Degtyarev, M. Y., Spiegel, A. M., and Jones, T. L. Palmitoylation of a g protein alpha i subunit requires membrane localization not myristoylation. *Journal of Biological Chemistry*, 269 (49):30898–30903, 1994.

DePaul, A. J., Thompson, E. J., Patel, S. S., Haldeman, K., and Sorin, E. J. Equilibrium conformational dynamics in an rna tetraloop from massively parallel molecular dynamics. *Nucleic Acids Research*, page gkq134, 2010.

Dessauer, C. W., Tesmer, J. J. G., Sprang, S. R., and Gilman, A. G. Identification of a giα binding site on type v adenylyl cyclase. *Journal of Biological Chemistry*, 273(40):25831–25839, 1998.

Deupi, X., Edwards, P., Singhal, A., Nickle, B., Oprian, D., Schertler, G., and Standfuss, J. Stabilized g protein binding site in the structure of constitutively active metarhodopsin-ii. *Proceedings of the National Academy of Sciences*, 109(1):119–124, 2012.

Dickson, C. J., Rosso, L., Betz, R. M., Walker, R. C., and Gould, I. R. Gafflipid: a general amber force field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter*, 8 (37):9617–9627, 2012.

Dirac, P. A. M. Note on exchange phenomena in the thomas atom. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 26, pages 376–385. Cambridge University Press, 1930.

Doemer, M., Maurer, P., Campomanes, P., Tavernelli, I., and Rothlisberger, U. Generalized qm/mm force matching approach applied to the 11-cis protonated schiff base chromophore of rhodopsin. *Journal of Chemical Theory and Computation*, 10(1):412–422, 2013.

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. Pdb2pqr: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research*, 35(suppl 2):W522–W525, 2007.

Doltsinis, N. L. and Marx, D. First principles molecular dynamics involving excited states and nonadiabatic transitions. *Journal of Theoretical and Computational Chemistry*, 1(02): 319–349, 2002.

Dror, R. O., Mildorf, T. J., Hilger, D., Manglik, A., Borhani, D. W., Arlow, D. H., Philippsen, A., Villanueva, N., Yang, Z., and Lerch, M. T. Structural basis for nucleotide exchange in heterotrimeric g proteins. *Science*, 348(6241):1361–1365, 2015.

Ebrey, T. G. [14] pka of the protonated schiff base of visual pigments. *Methods in Enzymology*, 315:196–207, 2000.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M., Eramian, D., Shen, M.-y., Pieper, U., and Sali, A. Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics*, pages 5–6, 2006.

Fahmy, K., Jäger, F., Beck, M., Zvyaga, T. A., Sakmar, T. P., and Siebert, F. Protonation states of membrane-embedded carboxylic acid groups in rhodopsin and metarhodopsin ii: a fourier-transform infrared spectroscopy study of site-directed mutants. *Proceedings of the National Academy of Sciences*, 90(21):10206–10210, 1993.

Field, M. J., Bash, P. A., and Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry*, 11(6):700–733, 1990.

Foreman, J. C., Johansen, T., and Gibb, A. J. *Textbook of receptor pharmacology*. CRC press, 2010.

Fotiadis, D., Jastrzebska, B., Philippsen, A., Müller, D. J., Palczewski, K., and Engel, A. Structure of the rhodopsin dimer: a working model for g-protein-coupled receptors. *Current Opinion in Structural Biology*, 16(2):252–259, 2006.

Frenkel, D. and Smit, B. Understanding molecular simulations: from algorithms to applications. *Academic, San Diego*, 2002.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. Gaussian09 revision d.01. Gaussian Inc. Wallingford CT 2009.

Galbiati, F., Guzzi, F., Magee, A., Milligan, G., and Parenti, M. Chemical inhibition of myristoylation of the g-protein gi1 alpha by 2-hydroxymyristate does not interfere with its palmitoylation or membrane association. evidence that palmitoylation, but not myristoylation, regulates membrane attachment. *Biochemical Journal*, 313:717–720, 1996.

Galés, C., Van Durm, J. J., Schaak, S., Pontier, S., Percherancier, Y., Audet, M., Paris, H., and Bouvier, M. Probing the activation-promoted structural rearrangements in preassembled receptor–g protein complexes. *Nature Structural & Molecular Biology*, 13(9):778–786, 2006.

Gat, Y. and Sheves, M. A mechanism for controlling the pka of the retinal protonated schiff base in retinal proteins. a study with model compounds. *Journal of the American Chemical Society*, 115(9):3772–3773, 1993.

Gieseking, R. L., Risko, C., and Brédas, J.-L. Distinguishing the effects of bond-length alternation versus bond-order alternation on the nonlinear optical properties of π-conjugated chromophores. *The Journal of Physical Chemistry Letters*, 6(12):2158–2162, 2015.

Glymour, C. and Scheines, R. Causal modeling with the tetrad program. *Synthese*, 68(1):37–63, 1986.

Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. H++: a server for estimating pkas and adding missing hydrogens to macromolecules. *Nucleic Acids Research*, 33(suppl 2):W368–W371, 2005.

Hall, M. A. *Correlation-based feature selection for machine learning*. PhD thesis, The University of Waikato, 1999.

Hamilton, N. Z. Correlation-based feature subset selection for machine learning. 1998.

Hess, B. P-lincs: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation*, 4(1):116–122, 2008.

Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

Hohenberg, P. and Kohn, W. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864, 1964.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

Hoyer, P. O., Hyvarinen, A., Scheines, R., Spirtes, P. L., Ramsey, J., Lacerda, G., and Shimizu, S. Causal discovery of linear acyclic models with arbitrary distributions. *arXiv preprint arXiv:1206.3260*, 2012.

Hug, S. J., Lewis, J. W., Einterz, C. M., Thorgeirsson, T. E., and Kliger, D. S. Nanosecond photolysis of rhodopsin: Evidence for a new blue-shifted intermediate. *Biochemistry*, 29(6): 1475–1485, 1990.

Humphrey, W., Dalke, A., and Schulten, K. Vmd: visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.

Hurley, J. H. Structure, mechanism, and regulation of mammalian adenylyl cyclase. *Journal of Biological Chemistry*, 274(12):7599–7602, 1999.

Jaeger, F., Fahmy, K., Sakmar, T. P., and Siebert, F. Identification of glutamic acid 113 as the schiff base proton acceptor in the metarhodopsin ii photointermediate of rhodopsin. *Biochemistry*, 33(36):10878–10882, 1994.

Jämbeck, J. P. and Lyubartsev, A. P. An extension and further validation of an all-atomistic force field for biological membranes. *Journal of Chemical Theory and Computation*, 8(8): 2938–2948, 2012.

Jastrzebska, B., Ringler, P., Lodowski, D. T., Moiseenkova-Bell, V., Golczak, M., Müller, S. A., Palczewski, K., and Engel, A. Rhodopsin–transducin heteropentamer: Three-dimensional structure and biochemical characterization. *Journal of Structural Biology*, 176(3):387–394, 2011.

Jastrzebska, B., Orban, T., Golczak, M., Engel, A., and Palczewski, K. Asymmetry of the rhodopsin dimer in complex with transducin. *The FASEB Journal*, 27(4):1572–1584, 2013.

Johnston, C. A. and Siderovski, D. P. Receptor-mediated activation of heterotrimeric g-proteins: current structural insights. *Molecular Pharmacology*, 72(2):219–230, 2007.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

Joung, I. S. and Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The Journal of Physical Chemistry B*, 112(30):9020–9041, 2008.

Kaya, A. I., Lokits, A. D., Gilbert, J. A., Iverson, T. M., Meiler, J., and Hamm, H. E. A conserved phenylalanine as a relay between the $\alpha5$ helix and the gdp binding region of heterotrimeric gi protein α subunit. *Journal of Biological Chemistry*, 289(35):24475–24487, 2014.

Kilts, J. D., Gerhardt, M. A., Richardson, M. D., Sreeram, G., Mackensen, G. B., Grocott, H. P., White, W. D., Davis, R. D., Newman, M. F., Reves, J. G., et al. $\beta2$-adrenergic and several other g protein–coupled receptors in human atrial membranes activate both gs and gi. *Circulation Research*, 87(8):705–709, 2000.

Kim, J. E., Tauber, M. J., and Mathies, R. A. Wavelength dependent cis-trans isomerization in vision. *Biochemistry*, 40(46):13774–13778, 2001.

Kohn, W. and Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133, 1965.

Lacerda, G., Spirtes, P. L., Ramsey, J., and Hoyer, P. O. Discovering cyclic causal models by independent components analysis. *arXiv preprint arXiv:1206.3273*, 2012.

Lagerström, M. C. and Schiöth, H. B. Structural diversity of g protein-coupled receptors and significance for drug discovery. *Nature Reviews Drug Discovery*, 7(4):339–357, 2008.

Laio, A., VandeVondele, J., and Rothlisberger, U. A hamiltonian electrostatic coupling scheme for hybrid car–parrinello molecular dynamics simulations. *The Journal of Chemical Physics*, 116(16):6941–6947, 2002a.

Laio, A., VandeVondele, J., and Rothlisberger, U. D-resp: Dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations. *The Journal of Physical Chemistry B*, 106(29):7300–7307, 2002b.

Laio, A., Gervasio, F. L., VandeVondele, J., Sulpizi, M., and Rothlisberger, U. A variational definition of electrostatic potential derived charges. *The Journal of Physical Chemistry B*, 108(23):7963–7968, 2004.

Lavery, R. personal communication.

Leach, A. R. *Molecular modelling: principles and applications*. Pearson education, 2001.

Lee, C., Yang, W., and Parr, R. G. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785, 1988.

Lemaître, V., Yeagle, P., and Watts, A. Molecular dynamics simulations of retinal in rhodopsin: from the dark-adapted state towards lumirhodopsin. *Biochemistry*, 44(38):12667–12680, 2005.

Lewis, J. W. and Kliger, D. S. Photointermediates of visual pigments. *Journal of Bioenergetics and Biomembranes*, 24(2):201–210, 1992.

Lewis, J. W., Szundi, I., Kazmi, M. A., Sakmar, T. P., and Kliger, D. S. Time-resolved photointermediate changes in rhodopsin glutamic acid 181 mutants. *Biochemistry*, 43(39):12614–12621, 2004.

Li, J., Edwards, P. C., Burghammer, M., Villa, C., and Schertler, G. F. Structure of bovine rhodopsin in a trigonal crystal form. *Journal of Molecular Biology*, 343(5):1409–1438, 2004.

Ligeti, E., Welti, S., and Scheffzek, K. Inhibition and termination of physiological responses by gtpase activating proteins. *Physiological Reviews*, 92(1):237–272, 2012.

Linder, M. E., Pang, I.-H., Duronio, R. J., Gordon, J. I., Sternweis, P. C., and Gilman, A. G. Lipid modifications of g protein subunits. myristoylation of go alpha increases its affinity for beta gamma. *Journal of Biological Chemistry*, 266(7):4654–4659, 1991.

Liu, Y., Kahn, R. A., and Prestegard, J. H. Structure and membrane interaction of myristoylated arf1. *Structure*, 17(1):79–87, 2009.

Liu, Y., Kahn, R. A., and Prestegard, J. H. Dynamic structure of membrane-anchored arf [bull] gtp. *Nature Structural & Molecular Biology*, 17(7):876–881, 2010.

Loisel, T. P., Ansanay, H., Adam, L., Marullo, S., Seifert, R., Lagacé, M., and Bouvier, M. Activation of the β2-adrenergic receptor-gαs complex leads to rapid depalmitoylation and inhibition of repalmitoylation of both the receptor and gαs. *Journal of Biological Chemistry*, 274(43):31014–31019, 1999.

López, C. S., Faza, O. N., Estévez, S. L., and de Lera, A. R. Computation of vertical excitation energies of retinal and analogs: scope and limitations. *Journal of Computational Chemistry*, 27(1):116–123, 2006.

Louet, M., Martinez, J., and Floquet, N. Gdp release preferentially occurs on the phosphate side in heterotrimeric g-proteins. *PLoS Computational Biology*, 8(7):e1002595, 2012.

Mahan, G. D. *Many-particle physics.* Springer Science & Business Media, 2013.

Markus, M., Andreas, R., Frank, W., Stefan, E., and Moritz, B. Dynamics of galphai1 interaction with type 5 adenylate cyclase reveal the molecular basis for high sensitivity of gi-mediated inhibition of camp production. *Biochemical Journal*, 454(3):515–523, 2013.

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Šali, A. Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure*, 29(1):291–325, 2000.

Marx, D. and Hutter, J. Ab initio molecular dynamics: Theory and implementation. *Modern Methods and Algorithms of Quantum Chemistry*, 1(301-449):141, 2000.

Maurer, P., Laio, A., Hugosson, H. W., Colombo, M. C., and Rothlisberger, U. Automated parametrization of biomolecular force fields from quantum mechanics/molecular mechanics (qm/mm) simulations through force matching. *Journal of Chemical Theory and Computation*, 3(2):628–639, 2007.

Meagher, K. L., Redman, L. T., and Carlson, H. A. Development of polyphosphate parameters for use with the amber force field. *Journal of Computational Chemistry*, 24(9):1016–1025, 2003.

Meijer, E. J. *Ab Initio Molecular Dynamics Preliminaries and Some Basic Concepts.* lecture notes, 2007.

Montoya, G., te Kaat, K., Moll, R., Schäfer, G., and Sinning, I. The crystal structure of the conserved gtpase of srp54 from the archaeon acidianus ambivalens and its comparison with related structures suggests a model for the srp–srp receptor complex. *Structure*, 8(5): 515–525, 2000.

Mou, T.-C., Masada, N., Cooper, D. M. F., and Sprang, S. R. Structural basis for inhibition of mammalian adenylyl cyclase by calcium. *Biochemistry*, 48(15):3387–3397, 2009.

Myers, J., Grothaus, G., Narayanan, S., and Onufriev, A. A simple clustering algorithm can be accurate enough for use in calculations of pks in macromolecules. *Proteins: Structure, Function, and Bioinformatics*, 63(4):928–938, 2006.

Nair, A. G., Gutierrez-Arenas, O., Eriksson, O., Vincent, P., and Kotaleski, J. H. Sensing positive versus negative reward signals through adenylyl cyclase-coupled gpcrs in direct and indirect pathway striatal medium spiny neurons. *The Journal of Neuroscience*, 35(41):14017–14030, 2015.

Nakamichi, H. and Okada, T. Crystallographic analysis of primary visual photochemistry. *Angewandte Chemie International Edition*, 45(26):4270–4273, 2006a.

Nakamichi, H. and Okada, T. Local peptide movement in the photoreaction intermediate of rhodopsin. *Proceedings of the National Academy of Sciences*, 103(34):12729–12734, 2006b.

Okada, T., Sugihara, M., Bondar, A.-N., Elstner, M., Entel, P., and Buss, V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 å crystal structure. *Journal of Molecular Biology*, 342(2):571–583, 2004.

Oldham, W. M. and Hamm, H. E. Heterotrimeric g protein activation by g-protein-coupled receptors. *Nature Reviews Molecular Cell Biology*, 9(1):60–71, 2008.

Pasqualato, S., Renault, L., and Cherfils, J. Arf, arl, arp and sar proteins: a family of gtp-binding proteins with a structural device for 'front–back'communication. *EMBO Reports*, 3(11): 1035–1041, 2002.

Patrick, G. L. *An introduction to medicinal chemistry.* Oxford University Press, 2013.

Pierre, S., Eschenhagen, T., Geisslinger, G., and Scholich, K. Capturing adenylyl cyclases as potential drug targets. *Nature Reviews Drug Discovery*, 8(4):321–335, 2009.

Preininger, A. M., Van Eps, N., Yu, N.-J., Medkova, M., Hubbell, W. L., and Hamm, H. E. The myristoylated amino terminus of gαi1 plays a critical role in the structure and function of gαi1 subunits in solution. *Biochemistry*, 42(26):7931–7941, 2003.

Preininger, A. M., Kaya, A. I., Gilbert III, J. A., Busenlehner, L. S., Armstrong, R. N., and Hamm, H. E. Myristoylation exerts direct and allosteric effects on gα conformation and dynamics in solution. *Biochemistry*, 51(9):1911–1924, 2012.

Rasmussen, S. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., et al. Crystal structure of the [bgr] 2 adrenergic receptor-gs protein complex. *Nature*, 477(7366):549–555, 2011.

Raw, A. S., Coleman, D. E., Gilman, A. G., and Sprang, S. R. Structural and biochemical characterization of the gtpγs-, gdp pi-, and gdp-bound forms of a gtpase-deficient gly42 → val mutant of giα1. *Biochemistry*, 36(50):15660–15669, 1997.

Reynolds, C. A., Essex, J. W., and Richards, W. G. Atomic charges for variable molecular conformations. *Journal of the American Chemical Society*, 114(23):9075–9079, 1992.

Ritter, S. L. and Hall, R. A. Fine-tuning of gpcr activity by receptor-interacting proteins. *Nature Reviews Molecular Cell Biology*, 10(12):819–830, 2009.

Robinson, G. A., Butcher, R. W., and Sutherland, E. W. cyclic amp. *Annual Review of Biochemistry*, 37(1):149–174, 1968.

Robnik-Šikonja, M. and Kononenko, I. An adaptation of relief for attribute estimation in regression. In *Machine Learning: Proceedings of the Fourteenth International Conference (ICML97)*, pages 296–304, 1997.

Röhrig, U. F., Guidoni, L., and Rothlisberger, U. Early steps of the intramolecular signal transduction in rhodopsin explored by molecular dynamics simulations. *Biochemistry*, 41 (35):10799–10809, 2002.

Röhrig, U. F., Guidoni, L., Laio, A., Frank, I., and Rothlisberger, U. A molecular spring for vision. *Journal of the American Chemical Society*, 126(47):15328–15329, 2004.

Röhrig, U. F., Guidoni, L., and Rothlisberger, U. Solvent and protein effects on the structure and dynamics of the rhodopsin chromophore. *ChemPhysChem*, 6(9):1836–1847, 2005.

Saam, J., Tajkhorshid, E., Hayashi, S., and Schulten, K. Molecular dynamics investigation of primary photoinduced eventsin the activation of rhodopsin. *Biophysical Journal*, 83(6): 3097–3112, 2002.

Sadana, R. and Dessauer, C. W. Physiological roles for g protein-regulated adenylyl cyclase isoforms: insights from knockout and overexpression studies. *Neurosignals*, 17(1):5–22, 2008.

Saey, T. H. Seeing the light: Transforming nerve cells into photoreceptors to restore vision. *Science News*, 187(11):22–26, 2015.

Sakmar, T. P., Franke, R. R., and Khorana, H. G. Glutamic acid-113 serves as the retinylidene schiff base counterion in bovine rhodopsin. *Proceedings of the National Academy of Sciences*, 86(21):8309–8313, 1989.

Sakmar, T. P., Franke, R. R., and Khorana, H. G. The role of the retinylidene schiff base counterion in rhodopsin in determining wavelength absorbance and schiff base pka. *Proceedings of the National Academy of Sciences*, 88(8):3079–3083, 1991.

Sandberg, M. N., Greco, J. A., Wagner, N. L., Amora, T. L., Ramos, L. A., Chen, M.-H., Knox, B. E., and Birge, R. R. Low-temperature trapping of photointermediates of the rhodopsin e181q mutant. *SOJ Biochemistry*, 1(1), 2014.

Schröter, G., Mann, D., Kötting, C., and Gerwert, K. Integration of fourier transform infrared spectroscopy, fluorescence spectroscopy, steady-state kinetics and molecular dynamics simulations of gαi1 distinguishes between the gtp hydrolysis and gdp release mechanism. *Journal of Biological Chemistry*, 290(28):17085–17095, 2015.

Seidel, R. D., Amor, J. C., Kahn, R. A., and Prestegard, J. H. Conformational changes in human arf1 on nucleotide exchange and deletion of membrane-binding elements. *Journal of Biological Chemistry*, 279(46):48307–48318, 2004.

Shatsky, M., Nussinov, R., and Wolfson, H. J. A method for simultaneous alignment of multiple protein structures. *Proteins: Structure, Function, and Bioinformatics*, 56(1):143–156, 2004.

Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(Oct):2003–2030, 2006.

Simon, M. I., Strathmann, M. P., and Gautam, N. Diversity of g proteins in signal transduction. *Science*, 252(5007):802–808, 1991.

Singh, U. C. and Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the ch3cl+ cl- exchange reaction and gas phase protonation of polyethers. *Journal of Computational Chemistry*, 7(6):718–730, 1986.

Singhal, A., Ostermaier, M. K., Vishnivetskiy, S. A., Panneels, V., Homan, K. T., Tesmer, J. J. G., Veprintsev, D., Deupi, X., Gurevich, V. V., Schertler, G. F. X., et al. Insights into congenital stationary night blindness based on the structure of g90d rhodopsin. *EMBO Reports*, 14(6): 520–526, 2013.

Singhal, A., Guo, Y., Matkovic, M., Schertler, G., Deupi, X., Yan, E. C., and Standfuss, J. Structural role of the t94i rhodopsin mutation in congenital stationary night blindness. *EMBO Reports*, page e201642671, 2016.

Smith, S. O. Structure and activation of the visual pigment rhodopsin. *Annual Review of Biophysics*, 39:309–328, 2010.

Sorin, E. J. and Pande, V. S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophysical Journal*, 88(4):2472–2493, 2005.

Spirtes, P., Glymour, C. N., and Scheines, R. *Causation, prediction, and search*. MIT press, 2000.

Standfuss, J., Edwards, P. C., D'Antona, A., Fransen, M., Xie, G., Oprian, D. D., and Schertler, G. F. X. The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature*, 471(7340):656–660, 2011.

Steinberg, G., Ottolenghi, M., and Sheves, M. pka of the protonated schiff base of bovine rhodopsin. a study with artificial pigments. *Biophysical Journal*, 64(5):1499, 1993.

Struts, A. V., Salgado, G. F. J., Martínez-Mayorga, K., and Brown, M. F. Retinal dynamics underlie its switch from inverse agonist to agonist during rhodopsin activation. *Nature Structural & Molecular Biology*, 18(3):392–394, 2011.

Sunahara, R. K. and Taussig, R. Isoforms of mammalian adenylyl cyclase: multiplicities of signaling. *Molecular Interventions*, 2(3):168, 2002.

Sunahara, R. K., Dessauer, C. W., and Gilman, A. G. Complexity and diversity of mammalian adenylyl cyclases. *Annual Review of Pharmacology and Toxicology*, 36(1):461–480, 1996.

Sunahara, R. K., Dessauer, C. W., Whisnant, R. E., Kleuss, C., and Gilman, A. G. Interaction of gsα with the cytosolic domains of mammalian adenylyl cyclase. *Journal of Biological Chemistry*, 272(35):22265–22271, 1997a.

Sunahara, R. K., Tesmer, J. J. G., Gilman, A. G., and Sprang, S. R. Crystal structure of the adenylyl cyclase activator gsα. *Science*, 278(5345):1943–1947, 1997b.

Taussig, R., Iniguez-Lluhi, J. A., and Gilman, A. G. Inhibition of adenylyl cyclase by gi alpha. *Science*, 261(5118):218–221, 1993.

Tesmer, J. J. G., Berman, D. M., Gilman, A. G., and Sprang, S. R. Structure of rgs4 bound to alf 4–activated g iα1: stabilization of the transition state for gtp hydrolysis. *Cell*, 89(2):251–261, 1997.

Tesmer, J. J., Sunahara, R. K., Johnson, R. A., Gosselin, G., Gilman, A. G., and Sprang, S. R. Two-metal-ion catalysis in adenylyl cyclase. *Science*, 285(5428):756–760, 1999.

Thompson, D. L. *Modern methods for multidimensional dynamics computations in chemistry*. World Scientific, 1998.

Thorgeirsson, T. E., Lewis, J. W., Wallace-Williams, S. E., and Kliger, D. S. Effects of temperature on rhodopsin photointermediates from lumirhodopsin to metarhodopsin ii. *Biochemistry*, 32(50):13861–13872, 1993.

Tully, J. C. Molecular dynamics with electronic transitions. *The Journal of Chemical Physics*, 93 (2):1061–1071, 1990.

Unni, S., Huang, Y., Hanson, R. M., Tobias, M., Krishnan, S., Li, W. W., Nielsen, J. E., and Baker, N. A. Web servers and services for electrostatics calculations with apbs and pdb2pqr. *Journal of Computational Chemistry*, 32(7):1488–1491, 2011.

van Keulen, S. C. and Rothlisberger, U. Effect of n-terminal myristoylation on the active conformation of gα_{i1}:gtp. *in publication*, 2016.

von Lilienfeld, O. A., Tavernelli, I., Rothlisberger, U., and Sebastiani, D. Variational optimization of effective atom centered potentials for molecular properties. *The Journal of Chemical Physics*, 122(1):014113, 2005.

Wall, M. A., Coleman, D. E., Lee, E., Iñiguez-Lluhi, J. A., Posner, B. A., Gilman, A. G., and Sprang, S. R. The structure of the g protein heterotrimer g iα1 β 1 γ 2. *Cell*, 83(6):1047–1058, 1995.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

Warshel, A. and Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249, 1976.

Wedegaertner, P. B. and Bourne, H. R. Activation and depalmitoylation of g sα. *Cell*, 77(7): 1063–1070, 1994.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106(3):765–784, 1984.

Wikipedia. Adenylyl cyclase. URL https://en.wikipedia.org/wiki/Adenylyl_cyclase.

Witten, I. H. and Frank, E. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

Wright, M. H., Heal, W. P., Mann, D. J., and Tate, E. W. Protein myristoylation in health and disease. *Journal of Chemical Biology*, 3(1):19–35, 2010.

Zhukovsky, E. A., Robinson, P. R., and Oprian, D. D. Changing the location of the schiff base counterion in rhodopsin. *Biochemistry*, 31(42):10400–10405, 1992.

Siri Camee van Keulen

EPFL SB ISIC LCBC
BCH 4122 (Batochime UNIL)
Av. F.-A. Forel 2
CH-1015 Lausanne
Switzerland

s.c.vankeulen@gmail.com

+41 78 867 1237

Education

PhD in Computational Biochemistry **02.2012—02.2017**
École Polytechnique Fédérale de Lausanne, Switzerland
Supervision: Prof. Ursula Röthlisberger

Dissertation: Computational investigation of intracellular signalling cascades, from G-protein-coupled receptors to adenylyl cyclase

Master in Chemistry, **cum laude** **09.2009—10.2011**
University of Amsterdam, the Netherlands
Supervision: Prof. Peter Bolhuis

My master project focused on investigating a part of the photo cycle of Photoactive Yellow Protein (PYP) using computer simulations (quantum mechanics/molecular mechanics molecular dynamics) and transition path sampling (TPS). During the project I wrote a code to apply TPS on the PYP problem.

Bachelor in Chemistry, **honours** student **09.2005—07.2009**
University of Amsterdam, the Netherlands

Current position

PhD researcher in Computational Biochemistry **02.2012—02.2017**
École Polytechnique Fédérale de Lausanne, Switzerland

In my first project I investigated via classical and quantum mechanics/molecular mechanics molecular dynamics simulations, the effect of light on the eye by studying the mechanism of rhodopsin's activation. In a second project, which is part of the Human Brain Project, I focused on how the activation signal of G-protein-coupled receptors, like rhodopsin, is transmitted in cells via the inhibitory G_i protein and how activation changes the conformation and the functionality of the G_i protein.

Publications

1. **Van Keulen, S. C.**, Gianti, E., Carnevale, V., Klein, M. L., Rothlisberger, U., Delemotte, L. (2016). *Does Proton Conduction in the Voltage-Gated H+ Channel hHv1 Involve Grotthuss-Like Hopping via Acidic Residues?*. The Journal of Physical Chemistry B.
2. **Van Keulen, S. C.**, Rothlisberger, U. (2017). "*Effect of N-terminal Myristoylation on the Active Conformation of* $G\alpha_{i1}$ *:GTP*". Biochemistry. 56(1): 271–280.

3. **Van Keulen, S. C.**, Rothlisberger, U., "*Exploring the Inhibition Mechanism of Adenylyl Cyclase Type 5 by N-terminal Myristoylated* $G\alpha_{i1}$*:GTP*". (under review for PLOS Computational Biology)
4. **Van Keulen, S. C.**, Solano, A., Rothlisberger, U., "*Exploration of the Deprotonation Pathway of Rhodopsin's Protonated Schiff Base* ". (to be submitted)
5. Nogly, P., Weinert, T., James, D., **Van Keulen, S. C.**, Liberatore, E., Bircher, M. P., Bhattacharyya, S, Rothlisberger, U., Standfuss, J. *Ultrafast Snapshot of Bacteriorhodopsin's Retinal Caught in Action at 10 ps.* (in preparation)

Talks

Invited talk "The effect of myristoylation on G proteins" at the Bioinformatique Structures et Interactions (BISI) laboratory in Lyon, France. March 2016.
Two talks at Human Brain Project seminar sessions in Barcelona, Spain, and in Julich, Germany. June and December 2014.

Teaching

Supervision of master thesis of Alicia Solano **09.2015—02.2016**
Thesis title: First steps towards an accurate and transferable force field for magnesium ions in proteins: inner-sphere ligand identification and absolute hardness analysis.

Co-wrote and **managed** the development of **course material** **09.2014—06.2015**

Teaching assistant and coordinator for two courses **09.2012—06.2015**
- Course title: Introduction to electronic structure methods (4 ECTS)
- Course title: Molecular dynamics and Monte Carlo simulations (2 ECTS)

Program experience with

Gromacs, Ambertools, CPMD, MODELLER, Gaussian, VMD, Python

Peer reviewed for

Journal of Chemical Theory and Computation
International Journal of Computational Biology and Drug Design
Physical Chemistry Chemical Physics

Nominated for

ISIC/EDCH Best Thesis Award 2017

Extra Curricular Activities

Organisation of scientific seminars in the Laboratory of Computational Chemistry and Biochemistry *École Polytechnique Fédérale de Lausanne, Switzerland*	**01.2014—present**
Member of the Curriculum Committee Chemistry *University of Amsterdam, the Netherlands*	**09.2009—08.2010**
Vice chairman of the Program Committee Chemistry *University of Amsterdam, the Netherlands*	**09.2008—08.2010**
Chairman of the student council of the faculty of Natural Sciences, Mathematics and Informatics *University of Amsterdam, the Netherlands*	**09.2007—08.2008**
Mentor of first-year Chemistry students *University of Amsterdam, the Netherlands*	**09.2006—08.2009**
Information officer for Chemistry Bachelor *University of Amsterdam, the Netherlands*	**09.2006—08.2007**
Chemistry tutor	**09.2005—08.2011**